

(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF

(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG



The present invention relates to regulatory DNA sequences containing promoter sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.

(57) Zusammenfassung

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen
Telomerase-Untereinheit und deren diagnostische und therapeutische Verwen-
dung

5 Aufbau und Funktion der Chromosomenenden

- Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten
- 10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).
- 15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.
- 20 Desweiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).
- 25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der
- 30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

15 Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

Das Enzym Telomerase synthetisiert die Telomere

20

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

25

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

30

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen
5 auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens
15 (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomerase-aktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder
20 ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34⁺38⁺-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischem Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt
25 werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgenese ist.

- 5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase
10 enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.
15

- Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweiligen Strahlen schädigen nicht nur die Tumorzellen,
20 sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle
25 Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden
30 dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapie von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-translationale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

25

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements“) be-

30

5
schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie vermitteln gewebs- oder zellspezifische Promotoreigenschaften.

10 Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15 Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20 Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25 Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

30 Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).
20

Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- α , IFN- β , IFN- γ , TNF, TNF- α , TNF- β , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 α , - β), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF, M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispielsweise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle β -Glukuronidase, pflanzliche β -Glukuronidase aus *Secale cereale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle β -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC), β -Galaktosidase (β -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH), β -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch
5 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch
10 andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen,
15 zen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

20 Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.
25

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

5 Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilsequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF β , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

20

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen
5 Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen,
15 die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der
20 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

- 5 A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reporter-gen mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reporter-genaktivität, um einen diagnostischen Wert zu erhalten;
- 10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reporter-gen-konstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assoziierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

25 A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Sprague Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

30 B: Zu Fig.1-A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

Fig. 2: Restriktionsanalyse der rekombinanten λ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene λ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des
10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei λ -Arme des Vektors EMBL3 Sp6/T7.

Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten λ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

15 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene λ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur
20 14: nicht verdaut.

25 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

30

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

10

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 5'GTAAAGTTGTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

15

20

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die dargestellte Sequenz entspricht SEQ ID NO 2.

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25

30

Exons sind als durchnummerierte schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymchnittstellen sind wie folgt gekennzeichnet: S, SacI; X, XhoI. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

10

15

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

20

25

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgespligte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

30

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

- 5 Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.
11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem
Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region
des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenz-
abschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts
sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.
- 10 Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.
Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-
flankierenden hTC Genregion schematisch dargestellt. Das ATG-
Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue
15 Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der
hTC Promotor-Luziferase Konstrukte dargestellt. Das promotorlose pGL2-
Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in
jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbil-
dungsseite sind ist die relative Luziferaseaktivität der verschiedenen
20 Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die
Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den
Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten
durchgeführt wurden.
- 25 Tab. 1: Exon Intron Übergänge des hTC-Gens
Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5'
Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor
und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke
unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Groß-
30 buchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind
gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in
10 Intron 2.

Tab. 1

3' Acceptor Sequence			5' Donor Sequence					
Intron	Exon	Exon No.	bp	Exon	Intron	Intron on No.	bp	
5' flanking Region								
caggcgcttccccgag	GTTTCAGGCAGCGCTGCGT	1	281	CGCCCTCCTTCCGCCAG	gtggcctccccgggtcg	1	104	
catgtccttctcgtttaag	GTGTCTGCTGGAAGGAGC	2	1354	TGGCTGCGCAGGAGCCAG	gtgaggagtggtggccgt	2	8616	
gaggggtctctctattgag	GGGTGGCTGTGTTCCGGC	3	196	TGCAAAAGCATTGGAATCAG	gtactgtatccccacgcca	3	2089	
cccatgtgtccccgag	ACAGCACTTGAAGAGGGTG	4	181	GTTCCGCAGAGAAAGAGG	gtggtgtgcttggttta	4	687	
ctgcctccactcacag	GCCGAGCTCTCACCTCGA	5	180	TGAGCTGTACTTTGTCAAG	gtgggtgccggggagacccc	5	494	
ccctctctctgcccggag	GTGGAATGACGGGGCGGT	6	156	CAAGGCCTTCAAGAGCCAC	gtaaggcttcaggtgtgata	6	>4660	180
ctcccgtctgcttccgag	GTCTTACCTTGACAGACC	7	96	TGCCGTGCTCATCGAGCAG	gtctgggcaactgcccctgca	7	980	
ctgtgtcttccccgccag	AGCTCCTCCCTGAATGAGG	8	86	CCGTGCGCATCAGGGGCAA	gtgagtcagggtggccaggt	8	2485	
gtattttcccttattttag	GTCTACTCTGCGTTTGGTG	9	114	CGGGATTGCGCGGGACGG	gtgaggcctcctcttcccc	9	1984	
cattgccccctctgacctag	GACCTGGTCCGAGGTGTC	10	72	ACGGAAACCTTCTCTCAG	gtgagcccggtcccggtg	10	1871	
attccccctgtgtctdag	CTATGCCCGGACCTCCATC	11	189	TGCAGAGCGACTACTCCAG	gtgagcgcaactggccgga	11	3805	
tctttcttggcgactctag	GTGAACAGCCTCCAGACGG	12	127	CCTGTTCTGGATTTCAG	gtgagcaggtgatggtca	12	880	
ctgtccgccaatcctctdag	GTTCACGCATGTGTCTG	13	62	TCCTGCTGCAGCGGTACAG	gtgagccgcccacaaagggg	13	3187	
agcctctgttttccccag	GGATGTGCTGGGGCCAA	14	125	CTGAAAGCCAAAGAACGAG	gtatgtgcagggtgcctggc	14	781	
tctgattttggccccgag	CCGAGACGCAGCTGAGTCG	15	138	CTGGGTCTCCTCAGGACAG	gcaagtgtgggtggagggc	15	536	
		16	664	TTTTTCAGTTTGTGAAAAA	3' flanking Region			

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Sp1	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschrte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5 Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca $1,5 \times 10^6$ Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa
10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml
15 Heringssperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit $1,5 \times 10^6$ cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte
20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.*
25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen
30 Experimenten jeweils 1 bis $1,5 \times 10^6$ Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmustert.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach λ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umklontiert (Beispiel 4).

Beispiel 3

Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde λ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).

Da die isolierte λ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

Beispiel 4

Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen XhoI und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes XhoI-BamHI-Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations-Restriktionsverdau mit den Enzymen XhoI und XbaI führte zur Klonierung von einem 6,5 kb großen XhoI-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem XhoI-Fragment.

Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym XhoI. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes XhoI-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes XhoI-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenklons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. XhoI. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie XhoI-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenklon P17 und 5' von Phagenklon P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers AP1 (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10

15

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen kloniert.

20

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTTG-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATAACC-3') zusammen mit der HDL-Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4 ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking
10 Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software (DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5 zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2 unterstrichen.

Contig1:

5 ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA 70
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGGG 140
 ACAAACCCAG AAATCAACAA CAGAGGAAT TTTGAAACT ATACAAACAC ATGAAATTA AACAAATATAC 210
 TTCTGAATGA CCAGTGAGTC AATGAAGAA TTAAGAGGA AATTGAAAA TTTATTTAAG CAAATGATAA 280
 CGGAACATA ACCTCTCAAA ACCCACGGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA 350
 AGCAGCTACA TCAAAAAAGT AGAAAAGCCA GCGCAGTGG CTCTGCCTG TAATCCGAGC ACTTTGGGAG 420
 10 GCCAAGGCGG GCAGATCGCC TGAGGTGAGG AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTCG 490
 CTACTAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG 560
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC 630
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT 700
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACTAAA CCTAAAAATT GTAAAGAAAA AGAAATAATA 770
 15 AAGATCAGAG CAGAAATAAA TGAAACTGAA AGATAACAT ACAAAAGATC AACAAAATTA AAAGTTGGTT 840
 TTTTGAAGAG ATAAACAAAA TTGACAAACC TTTGCCAGC CTAAGAAAA AGGAAAGAA ACCTAAATAA 910
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAAT CAAAGGATCA CTAGAGGCTA 980
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCCTAGA TGCATACCAAC 1050
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACA GACCAATAAC AATAATGGGA TTAAGCCAT 1120
 20 AATAAAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAAT GGCTTCCCTG CTGGATTTTA CCAATCATTT 1190
 AAAGAAGAAT GAATTCCTAAT CTACTCTAAA CTATTCGTA AAATAGAGGA AAGAACTACT CCAAACTCAT 1260
 TCTCATGGC CAGTATTACC CTGATTCCAA AACACAGACA AAACACATCA AAAACAAACA AACAACTCAT 1330
 CAGAAAGAAA GAAAACTACA GGCCAAATAT CCTGATGAAT ACTGATACAA AAATCCTCAA CAAACACATA 1400
 GCAAACCAAA TTAACACACA CCTTCGAAAG ATCATTCAAT GTGATCAAGT GGGATTATT CCAGGGATGG 1470
 25 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAAACTA 1540
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTTCTGCAC CTTCATGATA AAAACCCCTCA 1610
 AAAAACCCAG TATACAAAG ACATACAGCG CAGGCACAGT GGCTCACACC TGGCATCCCA GCACCTCTGG 1680
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGG ACTAGCCTGG GCAACAAAT GAGACCTGGT 1750
 CTACAAAAAA CTTTTTAAAA AAATTAGCCA GGCAATGATG CATATGCCTG TAGTCCAGC TAGTCTGGAG 1820
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGG TGCAGTGAGC CATGAACATG TCATGTGACT 1890
 30 CCAGCCTAGA CAACAGAA CAACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGAG 1960
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAG 2030
 AAGAAACATA TTTCAACATA ATAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGTAA 2100
 AGCCTTCTCT CTAAGATCTG GAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTCTAGA 2170
 35 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCATCCAAA CTGGAAAGGA AGAAGTCAA 2240
 TTTCTCTGTT TGCAGATGAT ATGATCTTAT ATCTGGAAAA GACTTAAGAC ACCACTAAAA AACTATTAGA 2310
 GCTGAATTTT GGTACAGCAG GATACAAAAT CAATGTACAA AAATCAGTAG TATTTCTATA TTCCAACAGC 2380
 AAACAATCTG AAAAAGAAAC CAAAAAGCA GCTACAAATA AAATTAACA GCTAGGAATT AACCAAGAA 2450
 40 GTGAAGATC TCTACATGA AAACATATAA ATGTGTATAA AAGAAATTGA AGAGGGCACA AAAAAGAGAA 2520
 AGATATTCCA TGTTTATAGA TTGGAAGAA AAATCTGTT AAATCTGTTA TACTACCCAA AGCAATTTAC 2590
 AAATTTCAATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT 2660
 TTGTACAGAA CCACAAAAGA CCCAGAAAG CCAAGCTAT CCTGACCAA AAGACAAAAG CTGGAAGCAT 2730
 CACATACCT GACTTCAAT TATACTACAA AGCTATAGTA ACCCAACTA CATGGTACTG GATAAATAAC 2800
 AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA 2870
 45 TTTTGAACA AGGTGCCAAG AACATACTTT GGGGAAAAA TAATCTCTTC AATAAATGGT GCTGGAGGAA 2940
 CTGGATATCC ATATGCAAAA TAACAATACT AGAATCTGT CTCTCACCAT ATACAAAAGC AAATCAAAAT 3010
 GGATGAAAGG CTTAAATCTA AAACCTCAA CTTTGAACCT ACTAAAAGAA AACACCGGAG AAATCTCCA 3080
 GGACATTGGA GTGGGCAAG ACTTCTGAG TAATTCCTG CAGGCACAGG CAACCAAGC AAAACAGAC 3150
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCAGCAA AGGAACAAT CAACAAGAG AGAGACAAC 3220
 50 CCACAGAATG GGAGAAATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA 3290
 GCTCAAACTA CTCTATAAGA AAAACACCTA ATAAAGCTGAT TTTCAAAAAT AAGCAAAAGA TCTGGGTAGA 3360
 CATTTCTCAA AATAAGTCAT ACAATGGCA AACAGGCATC TGAAATGTG CTCAACACCA CTGATCATCA 3430
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTCAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC 3500
 55 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAGGAA ACCCTGGAC ACTGTTGGTG GGAATGAAA 3570
 TTGCTACCAC TATGGAGAAC AGTTTGAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TACAGCAATC 3640
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGA TCAACAAGCT ATCTCCACTC CCACATTTAC 3710
 TGCAGCACTG TTATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAAAAG 3780
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA 3850
 CAGCATGGGG GGCACCTGGT AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT 3920
 60 CTCCTTACT TGTGGGAGCA AAAATTAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG 3990
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT 4060
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTAC CCTGATGTA 4130
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTATGTA TGCTATAGAT ATAAACCCCTA CTATATTTAA 4200
 AATTAAAAAT TTAATGGCCA GGCACGGTGG CTCTATGTCG TAATCCAGC ACTTTGGGAG GCCAGGGCGG 4270
 65 GTGGATCACC TGAGGTGAGG AGTTTGAAC CAGTCTGGCC ACCATGATGA AACCTGTCT TACTATAAGA 4340
 TACAAAAATT AGCCAGGCGT GGTGGCAGT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA 4410
 TTGCTTGAAC CTGGGAGCGG GAGGTTCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA 4480
 GAGCAAGACT CCATCTCAA ACAAAAACAA AAAAAAGAA ATTAATAATG TAATTTTTAT GTACCGTATA 4550
 AATATATACT CTACTATATT AGAAGTTAAA AATTAAACAA ATTATAAAG GTAATTAACC ACTTAATCTA 4620
 70 AAATAAGAAC AATGTATGTG GGGTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT 4690
 GTGAGGAGGG AACAGTGGAA GTTACTGTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAAAC 4760
 AAAGACAGGC TGGGAGAGT TAAAGAGCCA TTCTATAAGC CTAACAAACA CTGCTAATAA TGGTGAAAGG 4830
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCAGAGCTGC AGAATTGGCA CGTCTGATCA 4900
 CACCGTCTC TCATTACAGG TGCTTTTTT CTTGTGTGCT TGGAGATTT CGATTGTGTG TCGTGTGTTG 4970
 75 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC 5040
 TGGCAGGAAG CAGGTGGCTC TGTGACCTG AGCCACTTCA ATCTTCAAG GTCTCTGGCC AAGACCCAGG 5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCGATGAG	AAGCCTGCCT	5180
	CGTTGGTGAG	CAGCGCATGA	AGTGCCCTTA	TTTACGCTTT	GCAAAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGCGGCCAG	CGGGAATGCA	AGGAGTCAGA	AGCCTCCTGC	TCAAACCCAG	GCCAGCAGCT	ATGGCGCCCA	5320
5	CCCGGGCGTG	TGCCAGAGGG	AGAGGAGTCA	AGGCACCTCG	AAGTATGGCT	TAAATCTTTT	TTTACCTGA	5390
	AGCAGTGACC	AAGGTGTATT	CTGAGGGAAG	CTTGAGTTAG	GTGCCTTCTT	TAAACAGAA	AGTCATGGAA	5460
	GCACCCCTTCT	CAAGGGAAAA	CCAGACGCCC	GCTCTGCGGT	CATTACCTC	TTTCTCTCT	CCCTCTCTTG	5530
	CCCTCGCGGT	TTCTGATCGG	GACAGAGTGA	CCCCGTGGA	GCTTCTCCGA	GCCCGTGCTG	AGGACCTCT	5600
	TGCAAAAGGGC	TCCACAGACC	CCCGCCTGG	AGAGAGGAGT	CTGAGCCTGG	CTTAATAACA	AACCTGGATG	5670
10	TGGCTGGGGG	CGGACAGCGA	CGGCGGGATT	CAAAGACTTA	ATTCCATGAG	TAAATCAAC	CTTTCCACAT	5740
	CCGAATGGAT	TTGGATTTTA	TCTTAATATT	TTCTTAAATT	TCATCAAATA	ACATTACAGG	CTGCAGAAAT	5810
	CCAAAGGCGT	AAAACAGGAA	CTGAGCTATG	TTTGCCAAAG	TCCAAGGACT	TAATAACCAT	GTTCAGAGGG	5880
	ATTTTTCGCC	CTAAGTACTT	TTTATTGGTT	TTCAATAGGT	GGCTTAGGGT	GCAAGGGAAA	GTACACGAGG	5950
	AGAGGCGCTG	GCGGCAGGGC	TATGAGCAGC	GCAGGGCCAC	CGGGGAGAGA	GTCCCCGGCC	TGGGAGGCTG	6020
15	ACAGCAGGAC	CACCTGACCT	CCTCCCTGGG	AGCTGCCACA	TTGGGCAACG	GGAAGGCGGC	CAGCTCGCT	6090
	GTGACTCAGG	ACCCCATACC	GGCTTCTCTG	GCCCAACCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTGAAC	6160
	CCGTGGAAAC	GAACATGACC	CTTGCCCTG	TGCTTCCCTG	GGTGGGTCAA	GGGTAATGAA	TGGTGTGCA	6230
	GGAAATGGCC	ATGTAATAA	CAGGACTCTG	CTGATGGGGA	CCGTTCTCTC	CATCATTATT	CATCTTCACC	6300
	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAACTCTAG	TACAAACACC	6370
20	ACTCTTTTAC	TAGGCCACCA	GAGCAGGSC	CACACCCCTG	ATATATTAAG	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTACGCC	ACCAGGCTGG	GGTGACAACA	CGCGCTGAAC	AGTCTGTTC	TCTAGACTAG	TAGACCTG	6510
	CAGGCACTCC	CCCAGATTCT	AGGCGCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	GAGACTCAGC	6580
	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTG	TCCACACCCCT	CCGCCCTCCG	GCCTCAGCTT	CTCCAGCAGC	6650
	TTCTTAAACC	CTGGGTGGGC	CGTGTCCAG	CGCTACTGTC	TCACCTGTCC	CACCTGTGTC	TGCTCAGGC	6720
	ACGTAGCTCG	CACGGTTCCT	CCTCAGATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	CGCTGAAGT	6790
25	GAGGAGATTG	TGGCGCTCCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCCC	GATGACGGTT	6860
	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCCGTCTC	CTGTCTATCTG	CCGGGCGCTG	6930
	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGTGTC	TCTCTGCCCG	CTAGGGTCTC	7000
	GGGGTTTTTA	TAGGCATAGG	ACGGGGCGGT	GGTGGGCGAG	GGCGCTCTTG	GGAAATGCAA	CATTGGGGTG	7070
30	TGAAAGTAGG	AGTGCTCTGC	CTCACCTAGG	TCCACGGGCA	CAGGCGCTGG	GATGGAGCCC	CCGCCAGGGA	7140
	CCCGCCCTTC	TCTGCCACGC	ACTTCTCTGC	CCCCCTCCCT	CTGGAACACA	GAAGTGGCAGT	TCCACAAGC	7210
	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	GCCCCACAGC	CCTGGGAATT	7280
	ACCGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCGC	CTGTTTTATT	TTAATAGCTA	7350
	CAAGCAGGGG	AAATCCCTGC	TAAATGTGCC	TTTAAACAAAC	TGGTTAAACA	AACGGGTCCA	TCCGACCGGT	7420
35	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTC	ACCCCATGGC	7560
	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	CCTTTACTA	7630
	AAGCCAGTTT	CCTGGTTCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGATGGG	7700
	GGAAACCCGA	GGCTGTGCCA	TCTTTGCCAT	GCCCCAGTGT	CCTGGGCAGG	ATAATGCTCT	AGAGATGCC	7770
40	ACGTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCGCG	GCCCCAGGGC	CTTTGCAGGT	GTGATCTCCG	7840
	GAGGACCCCT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCGAAAA	GTAATCCAGG	GGTTCTGGGA	7910
	AGAGCGGGGC	AGGAGGGTCA	GAGGGGGGCA	GCCTCAGGAC	GATGGAGGCA	GTCACTCTGA	GGCTGAAAAG	7980
	GAGGGGAGGG	CCTCGAGCCC	AGGCCCTGCA	GCGCCTCCAG	AAGCTGGAAA	AAGCGGGGAA	GGGACCCCTC	8050
	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCTCC	8120
45	GTGCCATAGG	AGGCCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAAA	GCAACAGGAA	8190
	ACCCATGCAC	TGTGAATCTA	GGATTATTTT	AAAACAAAGG	TTTACAGAAA	CATCCAGGGA	CAGGGCTGAA	8260
	GTGCCCTCCG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	TACTTACTTT	8330
	CTGAGACAGA	GTTATGCTCT	TGTTGCCACG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	CTGCAACCTC	8400
	CGTCTCCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTCAGGC	GTGCACCACC	8470
50	ACACCCGGCT	AATTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	TCTCAAAATC	8540
	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	CACCTGCACCT	8610
	GGCCTATTTA	ACCATTTTAA	AACTTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAGGAGTT	CATGGAGTTC	8680
	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAATTC	TTCTGAGACT	GGGGATACAC	8750
	CGTCTCTTGA	CATATTACCA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	GGGCGAGTGG	8820
55	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	ATCAGGGCGC	8890
	AAGGTGGGAC	ACTGTCTCTG	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GTAGAAATTA	AAGTCCATCC	8960
	CTCTACTCT	ACTGGGATTG	AGCCCTTTC	CTATCCCCC	CCAGGGGCGG	AGGAGTTCTT	CTCACTCCTG	9030
	TGGAGGAAGG	AATGATACTT	TGTTATTTTT	CACTGCTGGT	ACTGAATCCA	CTGTTTCATT	TGTTGGTTTG	9100
	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACTC	TTGTGTCTCA	GGCTGGAGGG	AGTGCAATGG	CGCGATCTTG	9170
60	GCTTACTGCA	GCCTCTGCTC	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCGC	CTCCCATTTG	GCTGGGATTA	9240
	CAGGCACCCG	CCACCATGCC	CAGCTAATTT	TTTGATTTTT	TAGTAGAGAC	GGGGTGGGGT	GGGGTTCACC	9310
	ATGTTGGCCA	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCTCTC	AAAGTGCCTG	9380
	GATTACAGGT	GTAGGCCACC	ATGCCCACT	CAGAATTTAC	TCTGTTTAGA	AACATCTGGG	TCTGAGGTAG	9450
	GAAGCTCAC	CCACTCAAGT	GTTGTGGTGT	TTAAGCCAA	TGATAGAAAT	TTTTTATTGT	TGTTAGAACCA	9520
65	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGAGG	CGTTTCTCTG	CCATGCACAT	9660
	GGTGTTAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTTCT	TCTCTTCCCT	CTTTTAAAT	TGTGTTTTCT	9730
	ATGTTGGCTT	CTCTGCAGAG	AACAGTGTA	AGCTACAAC	TAACTTTTGT	TGGAACAAAT	TTTCCAAACC	9800
	GCCCTTTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTTAAAA	AGGCTTAGGG	ATCACTAAGG	9870
70	GGATTCTAG	AAGAGCGACC	TGTAATCTTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	GAGCGTGACA	9940
	GCCCGAGGAG	GGTGCAGGCG	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCTCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAGGTTATAC	TAAAGGTTGC	GTTTGTAGC	ATTTCAGTGT	TGCGCCACCT	CAGTACACG	10080
	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	AACCCGAGT	10150
	CTGGATTCTT	GGGAAGTCCT	CAGCTGTCTC	GCGGTTGTGC	CGGGGCCCA	GGTCTGGAGG	GGACCACTGG	10220
	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAAT	CGGCTCTCT	AGCTCTGAG	TCCGAGGCTT	GGAGCCAGGT	10290
75	GCCTGGACCC	CGAGGCTGCC	CTCCACCCTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	CATCTGCCAG	10360
	ACAGAGTGCC	GGGGCCGAGG	GTCAAGGCGC	TTGTGGCTGG	TGTGAGGCGC	CCGGTGGCGG	GCCAGCAGGA	10430
	GCGCTGGCT	CCATTTCCCA	CCCTTTCTCG	ACGGGACCGC	CCCGGTGGGT	GATTAACAGA	TTTGGGGTGG	10500

	TTTGCTCATG	GTGGGGACCC	CTCGCCGCT	GAGAACCTGC	AAAGAGAAAT	GACGGGCTG	TGTCAAGGAG	10570
	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACTCCGGGA	GGTCCCGCGT	GCCCGTCCAG	GGAGCAATGC	10640
	GTCTCGGGT	TCGTCCCCAG	CCGCGTCTAC	GCGCCTCCGT	CCTCCCTTC	ACGTCCGGCA	TTCTGGTGC	10710
	CCGGAGCCCG	ACGCCCCCG	TCCGGACTCG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	GCGGCCAAG	10780
5	GGTCGCCGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCCCTCCT	CGGGTTACCC	CACAGCCTAG	10850
	GCCGATTGGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCTT	GGGAGCGCGA	GCGCGCGCG	10920
	GGCGGGGAAG	CGCGGGCCAG	ACCCCGGGGT	CCGCGCGGAG	CAGCTGCGCT	GTCGGGGCCA	GGCCGGGCTC	10990
	CCAGTGGATT	CGCGGGCACA	GACGCCCAGG	ACCGCGCTCC	CCACGTGGCG	GAGGGAATCG	GGACCCGGGC	11060
	AGCCGTCTCG	CCCTTCACCC	TTCAGCTCC	GCCTCTCCG	GCGGAGCCCG	GCCCCGTCCC	GACCCCTCCC	11130
10	GGGTCCCGCG	CCCAGCCCCC	TCCGGGCCCT	CCCAGCCCCC	CCCTTCCTTT	TCCGGGGCCC	CGCCCTCTCC	11200
	TGCGGGCGCG	AGTTTCAGGC	AGCGCTGCGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	GGCCACCCCC	11270
	GCGATGCGCG	GCGCTCCCGG	CTGCGGAGCC	GTGCGCTCCC	TGCTGCGCAG	CCACTACCCG	GAGGTGCTGC	11340
	CGCTGGCCAC	GTTCTGTGCG	GCGCTGGGGC	CCCAGGGCTG	GCGGCTGGTG	CAGCGCGGGG	ACCCGGCGGC	11410
	TTTCGCGCGC	CTGGTGGCCC	AGTGCTGGT	GTGCGTGCCC	TGGGACGCGC	GGCCGCCCCC	CGCCGCCCCC	11480
15	TCCTTCCGCC	AGGTGGGCTT	CCCCGGGGTC	GGCGTCCGGC	TGGGGTTGAG	GCGCGCCGGG	GGGAACACAG	11550
	GACATGCGGA	GAGCAGCGCA	GGCGACTCAG	GGCGTTCCCC	CCGCAGGTGT	CCTGCCTGAA	GGAGCTGGTG	11620
	GCCCGAGTGC	TGCAGAGGCT	GTGCGAGCGC	GGCGGAAGA	ACGTGCTGGC	CTTCGGCTTC	GCGCTGCTGG	11690
	ACGGGGCCCC	CGGGGGCCCC	CCCCAGGCTT	TCACCACACG	CGTGCGCAGC	TACCTGCCCA	ACACGGTGAC	11760
	CGACGCACTG	CGGGGGAGCG	GGGCGTGGGG	GCTGCTGCTG	GCGCGCGTGG	GCGACGACGT	GCTGGTTTAC	11830
20	CTGCTGGCAC	GCTGCGCGCT	CTTTGTGCTG	GTGGCTCCCA	GCTGCGCTTA	CCAGGTGTGC	GGCGCGCGCG	11900
	TGTACCAAGT	CGCGCTGCCC	ACTCAGGCCC	GGCCCCCGCC	ACACGCTAGT	GGACCCCGAA	GGGCTCTGGG	11970
	ATGCGAAGCG	GCTTGAACCC	ATAGCTCAG	GGAGCGCGGG	GTCCCCCTGG	GCTTGCACGC	CCCCGGTGGC	12040
	AGGAGGCGCG	GGGCGAGTGC	CAGCCGAAGT	CTGCGGTTGC	CCAAGAGGGC	CAGGCGTGGC	GCTGCCCTTG	12110
25	AGCCGGAGCG	GACGCCCGTT	GGGCGAGGGT	CCTGGGCCCC	CCCCGGGCGG	ACGCGTGGAC	GAGTGACCGC	12180
	TGGTTTCTGT	GTGGTGTAC	CTGCCAGACC	CGCCGAAGAA	GCCACCTCTT	TGGAGGGTGC	GCTCTCTGGC	12250
	ACGCGCCACT	CCACCCATC	CGTGGGCGCG	CAGCACCACG	CAGGCCCCCG	ATCCACATCG	CGGCCACACC	12320
	GTCCCTGGGA	CACGCCCTGT	CCCCCGGTGT	ACGCCGAGAC	CAAGCACTTC	CTCTACTCTT	CAGGCGACAA	12390
	GGAGCAGCTG	CGGCCCTCCT	TCCTACTCAG	CTCTCTGAGG	CCACGCTGTA	CTGGCGCTCG	GAGGCTCGTG	12460
30	GAGACCATCT	TTCTGGGTTT	CAGGCCCTGG	ATGCCAGGGA	CTCCCCGCGC	GTTGCCCGCG	CTGCCCCAGC	12530
	GCTACTGGCA	AATGCGGCC	CTGTTTCTGG	AGCTGCTTGG	GAACACGCGC	CAGTGCCCCC	ACGGGGTGCT	12600
	CCTCAAGACG	CAGTGCCCGC	TGCGAGCTGC	GGTCAACCCA	GACGCGGGTG	TCTGTGCCCG	GGAGAACGCC	12670
	CAGGCGCTCT	TGGCGGGCCC	CGAGGAGGAG	GACACAGACC	CCCGTCCGCT	GCTGCAGCTG	CTCCGCCAGC	12740
	ACAGCAGCCC	CTGGCAGGTG	TACGGCTTCG	TGCGGGCCTG	CCTGCGCGCG	CTGGTGCCCC	CAGGCTCTGT	12810
35	GGGCTCCAGG	CACAACGAAC	GCCGCTTCTT	CAGGAACACC	AAGAAGTTCA	TCTCCCTGGG	GAGCATGTGC	12880
	AAGCTCTCGC	TGCAGGAGCT	GACGTGGAAG	ATGAGCGTGC	GGGACTGCGC	TTGGCTGCGC	AGGAGCCCAG	12950
	GTGAGGAGGT	GGTGGCGCTG	GAGGGCCAGC	GGCCAGAGGC	TGAATGCAGT	AGGGGCTCAG	AAAAGGGGGC	13020
	AGGCAGAGCC	CTGGTCTCTC	TGTCTCATC	GTACAGTGGG	CACACGTGGC	TTTTCGCTCA	GGACCTCGAG	13090
	TGGACACGGT	GATCTCTGCC	TCTGCTCTCC	CTCCTGTCCA	GTTTGATATA	ACTTACGAGG	TTACCTTCA	13160
40	CGTTTTGATG	GACACGCGGT	TTCCAGGCGC	CGAGGCCAGA	GCAGTGAACA	GAGGAGGCTG	GGCGCGGCAG	13230
	TGGAGCCGGG	TTGCCGGCAA	TGGGAGAGAG	TGCTTGAAG	CACAGACGCT	CTGGCGAGGG	TGCTGCGAGG	13300
	TTACCTATAA	TCCTCTTCGC	AATTTCAAGG	GTGGGAATGA	GAGGTGGGGA	CGAGAACCCC	CTCTTCTCTG	13370
	GGTGGGAGG	TAAGGGTTTT	GCAGGTGCAC	GTGGTCAGCC	AATATGCAGG	TTTGTGTTTA	AGATTTAATT	13440
	GTGTGTTGAC	GGCCAGGTGC	GGTGGCTCAC	GCCGGTAATC	CCAGCACTTT	GGGAAGCTGA	GGCAGGTGGA	13510
45	TCACCTGAGG	TCAGGAGTTT	GAGACGAGCC	TGACCAACAT	GGTGAACCCC	TATCTGTACT	AAAAATACAA	13580
	AAATTAGCTG	GGCATGTTGG	TGTGTGCCCT	TAATCCAGCG	TACTTGGGAG	GCTGAGGCAG	GAGAATCACT	13650
	TGAACCCAGG	AGGCGGAGCG	TGCAGTGAGC	TGAGATTGTG	CCATTGTACT	CCAGCCTGGG	CGACAAGAGT	13720
	GAAACTCTGT	CTTTAAAAAA	AAAAAGTGTT	CGTTGATTGT	GCCAGGACAG	GCTAGAGGGA	GGGAGATAAG	13790
	ACTGTTCTCC	AGCACAGATC	CTGGTCCCCAT	CTTTAGGTAT	GAAGAGGGCC	KCATGGGAGC	AGAGGACAGC	13860
50	AGATGGCTCC	ACCTGCTGAG	GAAGGGACAG	TGTTTGTGGG	TGTTACGGGG	ATGGTGCTCG	TGGGCCCTGC	13930
	CGTGTCCTCA	CCCTGTTTTT	CTGGATTGTA	TGTTGAGGAA	CCTCCGCTCC	AGCCCCCTTT	TGGCTCCAGC	14000
	TGCTCCACAG	CCCTACCGTG	GCAGCTAGAA	GAAGTCCCGA	TTTCAACCCC	TCCCCACAAA	CTCCCAAGAC	14070
	ATGTAAGACT	TCCGGCCATG	CAGACAAGGA	GGGTGACCTT	CTTGGGGCTC	TTTTTTTCTT	TTTTTTCTTT	14140
	TTATGGTGGC	AAAAGTCATA	TAACATGAGA	TTGGCACTCC	TAACACCGTT	TTCTGTGTAC	AGTGCAGAA	14210
55	TGCTAACTCG	GCGGTGTTTA	CAGCAGGTTG	CTTGAAATGC	TGCGTCTTGC	GTGACTGGAA	GTCCCTACCC	14280
	ATCGAACGGC	AGCTGCCTCA	CACCTGTGTC	GGCTCAGGTG	GACCACGCGC	AGTCAGATAA	CGGTCATGCA	14350
	ACCCAGTTTT	GCTTTTTTGT	CTCCAGCTTC	CTTCGTTGAG	GAGAGTTTGA	GTTCTCTGAT	CAGGACTCTG	14420
	CCTGTCAATT	CTGTCTCTCG	ACTTCAGATG	AGGTCACAAT	CTGCCCCTGG	CTTATGCAGG	GAGTGAGGCG	14490
	TGGTCCCCGG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGTTGCCCCC	AGGTGTCCCT	GTACCTGTA	14560
60	GGGTGAGTGA	GGCGGCGCCC	CCGGGTGTCC	CTGTCCCGTG	CAGCGTGATT	GAGGTGTGGC	CCCCGGGTGT	14630
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	ATCCCCGGGT	GTCCCTGTCA	CGTGTAGGGT	GAGTGAGGCG	14700
	TGCTCCCCGG	GTGTCCCTGT	CCCGTGCAGG	GTGAGTGAGG	CACGTGTCCC	GGGTGTCCCT	GTACAGTGCA	14770
	GGGTGAGTGA	GGCGGCGTCC	CCGGGTGTCC	CTCTCAGGTG	TAGGGTGAGT	GAGGCGCGGC	CCGAGGGTGT	14840
	CCCTGTACAG	TGTAGGGTGA	GTGAGGCGCC	GTCCCTGGGT	GTCCCTCCCA	GGTATAGGGT	GAGTGAGGCA	14910
65	CTGTCCCCGG	GTGTCCCTGT	CACGTGCAGG	GTGAGTGAGG	CGCGGCCCCC	GGGTGTCCCT	CTCAGGTGCA	14980
	GGGTGAGTGA	GGCGCTGTCC	CTGGGTGTCC	CTGTCTCGTG	TAGGGTGAGT	GAGGCTCTGT	CCCCAGGTGT	15050
	CCTTGGCGTT	TGCTCACTTG	AGCTTGTCTC	TGAATGTTTG	CTCTTTCTAT	AGCCACAGCT	GCGCGGGTTG	15120
	CCCATTTGCT	GGGTAGATGG	TGCAGGCGCA	GTGCTGGTCC	CCAAGCCTAT	CTTTTCTGAT	GCTCGGCTCT	15190
70	TCTTGGTCA	CTCTCCGTTT	CATTTTGTCTA	CGGGGACACG	GGACTGCAAG	CTCTCGCCCTC	CCGCGTGCCA	15260
	GGCACTGCAG	CCACAGCTTC	AGGTCCGCTT	GCCTCTGTTG	GGCCTGGCTT	GCTCACCACG	TGCCCCGCCAC	15330
	ATGCATGCTG	CCAATACTCC	TCTCCAGCT	TGTCTCATGC	CGAGGCTGGA	CTCTGGGCTG	CTGTGTCTG	15400
	CTGCCACGTG	TTGCTGGAGA	CATCCAGAA	AGGGTTCTCT	GTGCCCTGAA	GGAAAGCAAG	TCACCCACAG	15470
	CCCCCTCACT	GTCTGTTTTT	CTCCCAAGCT	GCCCTCTGCG	TTGGCCCCCT	TGGGTGGGTG	GCACAGCTTG	15540
	TCACCTTATT	CTGGGCACTT	CCCGCTCATT	GCTTAGGCTG	GGCTCTGCTT	CCAGTCGCCC	CCTCACATGG	15610
	ATTGACGTCC	AGCCACAGGT	TGGAGTGTCT	CTGTCTGTCT	CCTGCTCTGA	GACCCACGTG	GAGGGCGGCT	15680
75	GATCCCGCCA	GCCTTCGTCA	GACTTCCCTC	TTGGGTCTTA	GTTTTGAATT	TCACGTGATT	ACCTCTGACG	15750
	TTTCTATCTC	TCCATTGTAT	GCITTTTCTT	GGTTTATTCT	TTCACTTCTT	TTCTAGCTTC	TTAGTTTAGT	15820
	CATGCCCTTC	CCTCTAAGTG	CTGCCCTTACC	TGCACCTGT	GTTTTGTATG	GAAGTAATCT	CAACATCAGC	15890

	CACTTTCAAG	TGTTCTTAAA	ATACTTCAAA	GTGTTAATAC	TTCTTTTAAAG	TATTCCTTATT	CTGTGATTTT	15960
	TTTCTTTGTG	CACGCTGTGT	TTTGACGTGA	AATCATTITG	ATATCAGTGA	CTTTTAAAGTA	TTCTTTAGCT	16030
	TATTCCTGTGA	TTTCTTTGAG	CAGTGAGTTA	TTTGAACACT	GTTTATGTTT	AAGATATGTA	GAGTATCAAG	16100
5	ATACGTAGAG	TATTTTAAAGT	TATCATTITTA	TTATTGATTT	CTAAGTCACT	TGTGTAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTTGC	GGAGCCTTGC	TTTGTGATCT	AGTGTGTGCA	TGGTTTCCAG	AACTGTCCAT	16240
	TGTAAATTTG	ACATCCTGTC	AATAGTGGGC	ATGCATGTTT	ACTATATCCA	GCTTATTAAAG	GTCCAGTGCA	16310
	AAGCTTCTGT	CTCCTTCTAG	ATGCATGAAA	TTCCAAGAAAG	GAGGCCATAG	TCCCTCACCT	GGGGGATGGG	16380
	CTGTGTCATT	TCTTCTCGTT	TGGTAGCATT	TATGTAGGCG	ATTGTTAGGT	GCATGCACGT	GGTAGAATTT	16450
10	TTATCTTCCT	GATGAGTGAA	TCTTTTGGAG	ACTTCTATGT	CTCTAGTAAT	CTAGTAATTC	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTTCT	TTTGATTAGT	ATTTTCTGTC	TGTGTCTGTT	TTCTGCCCTT	16590
	AAITTTATATA	TATATATATA	TTTTTTTTTT	TTTTGAGACA	GAGTCTTGGT	CTGTCCGCCA	GGGTGAGTGC	16660
	AGTGGGTGTA	TCACAGGTCA	GTGTAACCTT	TACCTTCTGG	CCTGAGCCGT	CCTCTCACCT	CAGCCTCCTG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCCGTA	CACCTGGCTA	ATTTTTAAAT	TTTTTCTGGA	GACAGGGTCT	16800
	TGCTGTGTTG	CCCAGGCTGG	TCTCAAACTC	TTGGACTCAA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
15	CTGAATTACA	GGCATGAGCC	ACCATGTCTG	GCCTAATTTT	CAACACTTTT	ATATTCTTAT	AGTGTGGGTA	16940
	TGCTCTGTTA	ACAGCATGTA	GGTGAATTTT	CAATCCAGTC	TGACAGTCTG	TGTTTAACTG	GATAACCTGA	17010
	TTTATTTTCA	TTTTTTTGTG	ACTAGAGACC	CGCCTGGTGC	ACTCTGATTC	TCCACTTGCC	TGTTGCATGT	17080
	CCTCGTTTCC	TGTTTTCTCA	CCACCTCTTG	GGTTGCCATG	TGCGTTTCTT	GCCGAGGTGT	TGTTGATCCT	17150
	CTCGTTGCTT	CCTGGTCACT	GGGCATTTGC	TTTTATTCTT	CTTTGCTTAG	TGTTACCCCC	TGATCTTTTT	17220
20	ATTGTGCTTG	TTTGCTTTTG	TTTATTGAGA	CAGTCTCACT	CTGTACCCCA	GGCTGGAGTG	TGTTCCGCTG	17290
	ATCTCGGCTC	ACTGCAACCT	CTGCCCTCTC	GGTCAAGCA	GTCTCATTC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	CGCCCAACAC	CACGCCCTGC	TAATTTTTGT	ATTTTATAGT	GAGATAGGCT	TTCCACATGT	17430
	TGCCCAGGCT	GGTCTCAAA	TCTGACCTC	AAGTGATCTG	CCCGCCTTGG	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CCGGCATACC	TTGATCTTTT	AAAATGAAAGT	CTGAAACATT	GCTACCTTTG	17570
	TCTGAGCAA	TAAGACCTTT	AGTGATTTT	AGCTCTGGCC	ACCCCCCAGC	CTGTGTGCTG	TTTTCCCTGC	17640
	TGACTTAGTT	CTATCTCAGG	CATCTTGACA	CCCCCACAAG	CTAAGCATT	TTAATATTGT	TTTCCGTGTT	17710
	GAGTGTITCT	GTAGCTTTTG	CCCCGCCCTG	CTTTCTCTCC	TTTGTTCCTC	GTCTGTCTTC	TGCTCAGGCG	17780
	CCGCCGTCTG	GGGTCCCCTT	CCTTGTCTCT	TGCGTGGTTC	TTCTGTCTTG	TTATTGTCTG	TAAACCCAG	17850
	CTTTACCTGT	GCTGGCTTCC	ATGGCATCTA	GCGACGCTCG	GGGACCTCTG	CTTATGATGC	ACAGATGAGT	17920
30	ATGTGGAGAC	TCACGAGGAG	GGCGGTCTAT	TTGGCCCTGT	AGTGTCTGGA	GCACACAGTG	GCCAGCGTTC	17990
	CTTAGCCAGT	GAGTGACAGC	AACGTCCGCT	CGGCCCTGGT	TCAGCCTGGA	AAACCCAGG	CATGTCCGGG	18060
	TCTGGTGGCT	CCGCGGTGTC	GAGTTTGA	TCGCGCAAA	CTGCGGTGTC	GCGCCAGCTC	TGACCGTCTG	18130
	GCCTGGCGGG	GGAGTGTCTG	CTTCTCCCTT	TCTGCTTGGG	AACCCAGGCA	AAGGATGAGG	CTCCGAGCCG	18200
	TTGTCCGCCA	ACAGGAGCAT	GAGTGAGGCC	ATGTGGATA	TTTTAAATTT	TCTAGGCTGG	CGCGGTGGC	18270
35	TCACGCCCTG	AATCCACAGA	CTTTGGGAGG	CCAAGGCCGG	TGGATCACGA	GGTCAGGAGG	TGACAGCAT	18340
	CCTGGCCAAC	ATGATGAAC	CCCATCTGTA	CTAAACACAC	AAAAATTAGC	TGGCGTCTGT	GGCGGTGCC	18410
	TGTAAATCCA	GCTACTCGGG	AGGCTGAGGC	AGGAGAATTG	CTTGAACCTG	GGAGTTGGAA	GTTCAGATGA	18480
	GCCGACATTG	CACCACTGCA	CTCCAGCCTG	GCAACACAGC	GAGACTCTGT	CTCAAAAAAA	AAAAAAAAAA	18550
40	AAAAAATAAA	AATTTAGTA	GCCACATTA	AAAAGTAAAA	AAGAAAGGT	GAAATTAATG	TAATAATAGA	18620
	TTTTACTGAA	GCCAGCATG	TCCACACCTC	ATCATTTTAT	GGTGTATTG	GTGGGAGCAT	CACTCACAGG	18690
	ACATTTGACA	TTTTTTGAGC	TTTGTCTGCG	GGATCCCGTG	TGTAGGTCCC	GTGCGTGGCC	ATCTCGGCTC	18760
	GGACCTGCTG	GGCTTCCCAT	GGCCATGGCT	TTGTACCCAG	ATGGTGCAGG	TCCGGGATGA	GCTGCCAGG	18830
	CCCTCAGTGA	GCTGATGTG	CAGTGTCCGG	ATGGTGACAG	TCTGGGATGA	GGTCGCCAGG	CCCTGCTGTG	18900
45	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTACGGGGTG	AGGTCTCCAG	GCCCTCGGTG	AGCTGGAGGT	18970
	ATGGAGTCCG	GATGATGCAG	GTCCGGGGTG	AGGTCCGCCAG	GCCCTGCTGT	GAGCTGGATG	TGTGGTGTCT	19040
	GGATGGTGCA	GGTCAGGGGT	GAGGTCTCCA	GGCCCTCGGT	AAGCTGGAGG	TATGGAGTCC	GGATGATGCA	19110
	GGTCCGGGTT	GAGGTCCGCA	GGCCCTGCTG	TGAGCTGGAT	GTGTGGTGTG	TGGATGGTGC	AGGTCTGGGG	19180
	TGAGGTCAAC	AGGCCCTGCG	GTAGAGTGGG	TGTCCGGTGT	CTGGATGGTG	CAGGTCTGGA	GTGAGGTGCG	19250
	CAGACGGTGC	CAGACCATGC	GGTGAGCTGG	ATATGCCGGT	TCCGGATGGT	GCAGGTCTGG	GGTGAGGTTC	19320
50	CCAGCCCTTG	CTGTGAGTTG	GATGTGGGGT	GTCCGGATGC	TGCAGGTCCG	GTGTGAGGTC	ACCAGGCCCT	19390
	GCTGTGAGCT	GGATGTGTGG	TGTCTGGATG	GTGCAGGTCT	GGGGTGAAGG	TCCCCAGGCC	CTGCTTGTG	19460
	AGCTGGATGT	GTGGTGTCTG	GATGGTGCAG	GTCTGGAGTG	AGGTCCGCCAG	GCCCTCGGTG	AGCTGGATGT	19530
	GCAGTGTCCA	GATGGTGCAG	GTCCGGGGTG	AGGTCCGCCAG	ACCTTCCGGT	GAGCTGGATG	TCCGGTGTCT	19600
55	GGATGGTGCA	GGTCTGGAGT	GAGGTCCGCA	GGCCCTCGGT	GAGCTGGATG	TATGGAGTCC	GGATGGTGTG	19670
	GGTCCGGGGT	GAGGTCCGCA	GACCTGCTGT	TGAGCTGGAT	GTGCGGTGTC	TGGATGGTAC	AGGTCTGGAG	19740
	TGAGGTCCGC	AGACCTGCT	GTGAGCTGGA	TATGCCGTGT	CCGGATGGTG	CAGGTCCAGG	GTGAGGTCTC	19810
	CAGGCCCTCG	GTGAGCTGGA	GGTATGGAGT	CCGGATGATG	CAGGTCCGGG	GTGAGGTCCG	CAGGCCCTGC	19880
	TGTGAACCTG	ATGTCCGGCG	TCTGGATGGT	GCAGGTCTGG	GGTGTGGTGC	CCAGGCCCTC	GGTCAGTCTG	19950
60	AGGTATGGAG	TCCGGATGAT	GCAGGTCCGG	GGTGAGGTGC	CCAGGCCCTG	CTGTGAGCTG	GATGTCCGGC	20020
	GTCTGGATGG	TGCAGGTCTG	GGGTGTGGTC	GCCAGGCCCT	CGGTGAGCTG	GAGGTATGGA	GTCCGGATGA	20090
	TGCAGGTCCG	GGGTGAGGTT	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTCCAGTCCG	20160
	GGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCTG	TATCCGGATG	GTGAGGTCTC	GGGTGAGGTC	20230
	CACCAAGGCC	TGCGGTGAGC	TGGTTGTGCG	GTGTCCGGTT	GCTGCAGGTC	CGGGGTGAGT	TCCCAAGGCC	20300
65	CTCGGTGAGC	TGGATGTGCG	GTGTCCCGCT	GTCCGGATGG	TGCAGGTCCA	GGGTGAGGTC	GCTAGGCCCT	20370
	TGGTGGGCTG	GATGTGCCGT	GTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	20440
	GATGTGCCGT	GTCTGCATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGGGCTG	GATGTGTGGT	20510
	GTCCGGATGG	TGCAGGTCCG	CGGTGAGGTC	GCCAGGCCCT	GCTGTGAGCT	GGATGTGCGG	TGTCTGGATG	20580
	GTGCAGGTCC	GGGTGAGGTC	AGCCAAGGCC	TTCGGTGAGC	TGGATGTGGG	GTGTCCGGAT	GGTGCAGGTC	20650
70	CGGGGTGAGG	TCCCAAGGCC	CTCGGTGAGC	GTGATATGTC	GGGTGCCGGA	TGGTGCAGGT	CCGGGTGAGG	20720
	GCTACCAAGC	CCTCGGTTA	GCTGGATGTC	CGGTGTCTGG	ATGGTGACAG	TCCGGGGTGA	GGTCCCGAGG	20790
	CCCTGCTGTG	AGCTGGATGT	GCTGTATCCG	GATGGTGCAG	GTCCGGGGTG	AGGTCCGCCAG	GCCCTGCAGT	20860
	GAGCTGGATG	TGCTGTATCC	GGATGGTGCA	GGTCTGGCGT	GAGGTGCCCA	GGCCCTGCGG	TTAGCTGGAT	20930
	ATGCGGTGTC	GGATGGTGCA	GGTCCGGGGT	GAGGTACCCA	GGCCCTGCGG	TTAGCTGGAT	GTGCGGTGTC	21000
	CGGATGGTGC	AGGTCTGGGG	TGAGGTCCGC	AGGCCCTGCT	GTGAGCTGGA	TGTGTGTAT	CCGGATGGTG	21070
75	CAGGTCCGGG	GTGAGGTCCG	CAGGCCCTGC	GGTGTGCTGG	ATGTGCTGTA	TCCGGATGGT	CGAGGTCTGG	21140
	CGTGAGGTCC	CCAGGCCCTG	CGGTGAGCTG	GATGTGCAGT	GTACAGGTGG	TGCAGGTCCG	GGGTGAGGTC	21210
	GCCAGGCCCT	GCGGTGGGCT	GTATGTGTGT	TGTCTGGATG	GTGCAGGTCC	GGGTGAGTTC	GCCAGGCCCT	21280

TCGCGGTGAGC TGGATGTGTG GTGTCTGGAT GCTGCAGGTC CGGGGTGAGT TCGCCAGGCC CTCGGTGAGC 21350
 TGGATATGCG GTGTGCCCGT GTCCGAATGG TGGAGTGTCCA GGGTGAGGCT CGCCAGGCGT TGGTGGGCTG 21420
 GATGTGCCGT TGC CGGATGTTG TCGAGGTCTG GGGTGAGGTC ACCAGGCCCT GGGTGAGGCT GATGTGCCGT 21490
 CTCGGGATGG TCGAGGTCTG GGGTGAGGTC ACCAGGCCCT GGGTGAGGCT GATGTGCCGT 21560
 TTTAAGGGGT TGGCTGTGTT CGCGGCCAG AGCAAGCTGT GCGTGAGGAG ATCTCGGCCA AGTCTCTGCA 21630
 CTGGCTGATG AGTGTGTACG TCGTTCGAGCT GCTCAGGTCT TCTCTTTATG TCACGGAGAC CACGTTTCAA 21700
 AAGAACAGGC TCTTTTTCTA TCCGGAAGCT GTCTGGAGCA AGTTGGCAAAG CATTTGGAATC AGGTGATGTA 21770
 TCCCCAGGCC AGGCCCTCTG TCTTCAAGT CTTGGAAAC CAGCCCGGCC TACAGTATCG CCGTCTCTCA 21840
 CTGGCTGTG TCTCCCTGGC TGTGTCAGCT TGGGCTGGGA GCGGAGGGCC CCGCTCACGG CCGTGGTCAA 21910
 GTGGATTCTG TCGAAGGCTG TGACTGCCTG GAGCTCACGT TCTCTTACT TGTAAATCAG GAGTTGTGTC 21980
 CAAGTGGTCT CTAGGTTTTG TAAAGCAGAA GGGATTTAAA TTAGATGGAA ACACATCCAC TAGCTCCTT 22050
 GCGTTTCCCT GGGATGTGGG TCTGATCTCT TCTCTCTTT TTTTCTCTT TTTGAGATCT AGCTCTACTC 22120
 TGTTCGCCAG CTGGAGATG AGTGGCATRA TCTTGGCTCA CTGCAACCTC CACCTCTCTG GTTTAAGCGA 22190
 TCTCAGGAGC TCAGCCCTCTT AAGTAGCTGG GATTACAGGC ACCTCGCCAC AGCCCTGGCT AATTTTTGTA 22260
 CTTTTAGGAG AGACGGGGTG TACCATGTT GCGCAGGCT GTCTCGAAT CATGACCTCA GGGTATGCCA 22330
 CCACCTTTGG CTCCCAAAGT TCGGCTGTTA CAGGCTAAGC CACCGTGCCC AGCCCCCGAT TCTCTTTTAA 22400
 TTTATGCTGT TCTGTATGAA TCTTCAATCT ATTTGGATTA GGTGATGAAA GGATAAAAT CACCCACTAT 22470
 GGGCATCTAC TGCAGGGAGC ACCTGTGCAG GGAGCACCTG GGGATAGGAG AGTTCCACCA TGAGCTAACT 22540
 TCTAGGTGGC TGCATTTGAA TGGCTGTGAG ATTTGTCTG CGATGTCTCG CTGATGAGAG TGTGAGATTTG 22610
 TGACAGTAA AGCTTGGATT TGCATCAGTG AGGGAGCGGA CGCGTGGTCT GGGAGATGCC AGGCTGGCTG 22680
 AGGCCAGGCC ATGGTATTAG CTCTCCCGTG TCCCGCCGAG GCGTACTGTG GAGGGCTTTA TGCAGAAAGT 22750
 CAGGGCTTCC CAGGCTCCCC TGACACATCG AGTCCCTGGG GGGCTTTGTG ACACCCCATG CCCCAAATCA 22820
 GGAATGCTCG AGAGGGAGCT GAGCAGACAG CTCTGTAGAG GTAAACAGC GTTGGGAACT GGGAGCCCCA 22890
 CTGGTGCTG GGGCCATTCT TTTGATCTG GGGGAGGGTC CTTGCTTTCC CTGTGGGAGC AGGTTAATAC 22960
 ACAATGACCC TTACTTAGAC TTTACACGTA TTTAATGGTG TCGCAGCCAA CATGGTCTAT TGACCAAGTAT 23030
 TTTGGAAGAA ATTTAATTTG GGTGACCGGA AGGAGCAGAC AGACGTGGTG GTCCCAAGAT TGTCTCTTGT 23100
 CACTACTGGG ACTGTTTGTG TCTTTGGAGG CCGCTCTCTC CTGGACAGGG TACGCTGCC 23170
 TTTTACTCT CTGTGGCCTG CCGCTCTGCG TCAAGGCCACC AGCTCCGGAG ACACCCGGCC CCGAGTGTCC 23240
 ACCGAGTGCC AGGCTGTACG CCACAGATCG CGATGGCCAG GTGTGGCCGC TCACGCCCCC TGCCGCCCAT 23310
 GGGTGTTTT GGGGAAAGG GCCAAGGGCA GAGGTGTGAC GAGACTGGTG GGGTCTAGAG AGCTGATTCT 23380
 GCTCCTTGGC TGAGCTGCC TACGACGCTC CTCCCGCCCT CTCCATCTGA AGGATGTGG CTCITTTTAC 23450
 TCGGGGGTCC TGCCTGGGGC GAGCCTTGGG TACCCCCAGT TACCTGTACCA GAGGAGCAGG CATCTCTGTG 23520
 GGGAGGGCAT GGGTTACGT GCGCCCAAGT CGACGCTGGG ACCAGGCTCC CTGGTGCTGA TGGTGGGACA 23590
 TGACCCCTGG GGGTTGACG CCGGCACTGG CCGTCCAGG GTTGACTATA GGACACAGTG TCCAGGTGCC 23660
 CTGCAAGTAG AGGGGCTCTC AGAGGCGCTC GCGTGGCATG GGTGGACGTG GCGCCGGGCA TGGCCTTACG 23730
 CTGTGCTGC CCGGGTGGC CTGACCCCTT TGGGGCTTG TGGGCTCCG TGTGCTCCG TGAGCTTCCC 23800
 CCTAGTCTGT TGTCTGGCTG AGCAAGCCCTC CTGAGGGGCT CTCTATTGCA GACAGCACTT GAAGAGGGTG 23870
 CAGCTGCGGG AGCTGTCCGA AGCAGAGGCT AGGACAGCAT GGGAGACGAG GCGCCCGCTG CTGACGTCCA 23940
 GACTCCGCT CAGTCCCCAAG CTTGACGGCG TCGCGCCATG TGTGAACATG GACTAGCTG TGGGAGGACA 24010
 AAGCTTCCG CAGGAAAAGG GGGTGGCTGT GCTTTGTGTT AACTTCTCTT TTAACACAGG GTGCGTTTGA 24080
 GCGCCACAT TGGATACAG TTAGTGAAG GCGCCGAGG AGGGGCCGAC GGCACAGAAA AGGGCCATGG 24150
 CACGGCCGCA ACCATTGTTG GCGCACAGT AGGTGGCCGA GGTGCGGTG CTTCCAGAAA AGCAGCGTGG 24220
 GGGGTGAGG GGGATCTCTG GGGCAGGAG AGGCTCTGAG GACCCAAGA AGCAGCCGGG CACAGGCGCTG 24290
 GATGACGAG GCGCCGAGCT CTTGATTCG TGTCTGCTG TGTGCGCAT GGGCTCTGAG GCTCTCTGAG 24360
 CGGGGCGCCG GCGCAGGCC AGCATGTGCA GGAGCCCAAC GGGCTCTGAG GCTCTCTGAG CTCTCTGAG 24430
 GGGCTCTGCA CCGCCCCCTT GTGGCTGGG TGGCTCGGT GACCCCTCA TCTGAGGAGA GTGTGGGGTG 24500
 AGGTGAGAC AGGTGTGGCA TGAGGATCTG TGTGCAACA CACATCGCGC CAGGAAGGCT TTTCAACAGC 24570
 GGTCTAGGA AGCTGGGAGG GTTCTAGGT CCGGGTCTG GGTGGCTGG GGAACCTGGG AGGGGCTGCT 24640
 TCTCCCTG GTCCTATGG TGGGTGGGC ACTTGGCCG ATCCACTTT CTGACTGTCT CCCATGCTGT 24710
 CCGCCGAGG CCGAGCGCTC CACCTCGAGG TGAAGGCAC TGTTCAGGCT GCTCAACTCT GAGCGGGCCG 24780
 GCGCGCCGG CCGCTTGGG CTCTCTGTG TGGGCTTGA GCGGCTTGA GATATTCAT AGGCGCTTGG CACACTTCTG 24850
 GGTGCGTGT GCGGCGCAGG ACCCGCCGCT TGAGCTGTAC TTTGTCAAG TGGGTGGGCG GAGACCCCTG 24920
 GAGGAGCCCT CTGTGACATT GGAAGTGCT CCGCTGATTG CACCTCAT TGGTGGGAG AGGTACTCT 24990
 GGGTGGGCG CAGGAGATGC AGTGTGCCCT GTCACTTTG AGGACACACC TGGCCTATT GGTGAGGCC 25060
 TTCAGCTTT TCTGCAGCAC ATGGGCGCGA TGTGACACC TGACTGCCG GCGTCTATT CCGAAGGAG 25130
 TCCCACTGG ATTCCAGTT CCGTCAGAGA AGGAACCGCA ACGGCTCAGC CACCAGGCC CCGTGCTTG 25200
 CACCCCAAG TGTAGCCAGG GGTCTCTGT CCTGAGGCTC AGAGAGGGGA CACAGCCCGC CTGCGACTTG 25270
 GGGTCTGAG TGTGGGGGT GAGCAGAGA GTGGGGGACA CCGCCAGGCC AGGCCCTGAG CGGAGAGGTG 25340
 ATGTCTGAG TTTCTGCTGG CCACTGTGAC TCTCTGCGC TCCACTACCA CAGAGTGGAT TGACGGGGCG 25410
 TAGGACACAT ATCCCCAGG CACAGGCTAC GAGGATCAT CCAGCATCA TCAAAACCCA AACAACGAT 25480
 TGCCTGCTG GGTATGCCGT GGTCCAGAAG CCGGCCCATG GGCAGCTCG CTCAAGCTTCA GAATGCCAGT 25550
 TAAGGTTTAC GTGTGATGT TGTGTCCAGG ATGTGTGCT ACTGTGCTCT TTTACATAT TTTGATATCA GTGTGTGGA 25620
 GTGTCTGTA TCGGTTTCTG TGTGTTCTG TATCTGTGGC GTGTGTCTG GTGACAGTG CATGTCTAT AGCTGTGTG 25690
 CCGTGTGTC GTGGTGCATG GTGTGCTGT GTGTGTCTG TGTGTGCTG TGTGCTGAT CTATGGCAT 25760
 CCATGGTGTG GTGTGCTGT TGTGTGCTG TGTGTGCTG TGTGTGCTG TGTGTGCTG TGTGTGCTG 25830
 CATGTCTGTG ATGTGCTAT TGTGTGCTG TGTGTGCTG TGTGTGCTG TGTGTGCTG TGTGTGCTG 25900
 GGTGTGTGTC GCGCTTGGC TTTACTCTG CTCTCTCAG CATGTGTCG GGTGTGTCG TCTACGCTT 25970
 CCGGTGCTG TTTGGGAGC TCCACATTCA TTTGAGGAGG ATTAGGAGG TCTGTGCTG TCTGTGCTG 26040
 GCGGTGGGCC TTTGGAGACT TAAACCGAGT TTGAGAGGAG ATTAGGAGG CCGTGTGCTG TCTGTGCTG 26110
 CCGTGGACCC CCGAGACCC CAGCTGTGCC TATGCGCGCT TATGCGCGCT CCGTGTGCTG TCTGTGCTG 26180
 TCGCTCCCCG GGACACACT CTCCAGAGC GCGCGGGGCG CTTGCGGCTG TCTGTGCTG 26250
 TGGGCTTGG TTTCCACCA GTGGCTATGA GCACGCTGA GGGGTAGCC TCTAAAGTCT TCGCAGGCC 26320
 GGGTGCAGAG GTGAAGAA ATCCCTGGAG CTTGCGTCTG GGGAGAGCA CATGTGGAAA CCCACAGGA 26390
 CCTCTTTCTC TGACTTCTTG AGCT

Contig 2:

	TGTGGGATTG	GTTTTCATGT	GTGGGATAGG	TGGGGATCTG	TGGGATTGGT	TTTTATGAGT	GGGGTAACAC	70
	AGAGTTCAAG	GCGAGCTTTC	TTCTCTGATG	GGGTCTGCAG	GTGCTCCAAC	AGCTTTATTG	AGGAGACCAT	140
5	ATCTTCCTTT	GAACATAGGT	CGGGTTTATA	GTAAGTCAGG	GGTGTGGAGG	CCTCCCCCTGG	GCTCCCTGTT	210
	CTGTTTCTTC	CACCTCTGGG	TCGTGTGGTG	CCTGCTGTGG	TGTGTGGCCG	GTGGGCGAGG	CTTCCAGGCC	280
	TCCTTGTGTT	CATTGGCCCTG	GATGTGGCCC	TGGCTACGCT	CCGTCTTTGG	AATTCGCCCTG	CGAGTTGGAG	350
	GCTTTCCTTC	TTTTCTTTTT	TCCTTCTTTT	TTTTTTTTTT	TGATAACAGA	GTCTCGCTCT	TTTTTGCCCA	420
10	GGCTGGAGTG	GTTTGGCGTG	ATCTTGGCTC	ACTGCAACCT	GTGCTTCTCTG	AGTTCAAGCA	ATTCTCTTGC	490
	CTCAGCCTCC	CAAGTAGCTG	GAATTATAGG	CGCCCAACCAC	CATGCTGACT	AATTTTTGTA	ATTTTAGTAG	560
	AGACGAGGTT	TCTCCATGTT	GGCCAGGCTG	GTCTCGAACT	CCTGACCTCA	GGTGATCCTC	CCACCTCGGC	630
	CTCCCAAAGT	GCTGGGATGA	CAGGTGTGAA	CCGCCGCGCC	CGGCCGAGAC	TCGCTTCTCTG	CAGCTTCCGT	700
	GAGATCTGCA	GCGATAGCTG	CCTGCAGCCT	TGGTGTCTGAC	AACCTCCGTT	TTCTTCTCTC	AGGTCTCGCT	770
	AGGGGTCTTT	CCATTTCATG	ACTCTCTTCA	CAGAAGAGTT	TCACGTGTGC	TGATTTCCCG	GCTGTTTCTC	840
15	CGCTAATTGG	TGTCTGCTGT	TTATCGATGG	CCTCCTTCCA	TTTCTTTTAG	GCTTTGTTTA	TTGTTGTTTT	910
	TCCGGCTCCT	TGAAGGAAAA	GTTTCGATTA	TGGATGTTTG	AACCTTCTTT	TCTAAACAAG	CATCTGAAGT	980
	TGCCGTTTTT	CCTCTAAAGC	AGGGATCCCG	AGGCCCTTGG	CTGTGGAGTG	GCACCGGTCT	GGGGCTGTT	1050
	AGGAACCCGG	CGCACAGCGG	GAGGCTAGGT	GGGGTGTGGG	GAGCCAGCGT	TCCCGCCTGA	GGCCCGCCCC	1120
	TCTCAGATCA	GCAAGTGGCT	CGGGTGCTCA	GAGGCGCACA	CACCCCTACTG	AGAAGTGTGC	GTGAGAGGGG	1190
20	TCTAGATTCT	GTGCTCCTTA	TGGGAATCTA	ATGCCCTGATG	ATCTGAGGTG	GAACCGTTTG	CTCCCAAAAC	1260
	CATGCCCTTC	CCCACTGCTG	TCCTGTGGAA	AAATCGTCTT	CCACGAAACC	AGTCCCTGGT	ACCACATATG	1330
	TGGGGGACCC	TGTGCTAAAG	ACCTGCTTCA	GCAGCCTCTC	GTCAAGTGTG	ATATATTGGC	TTTTCTGTGT	1400
	TGAGTCCAGA	ATAATTACGG	ATTCTCTGTA	TGCTTCCCGC	CGACCTCAGA	CCCATGGGCT	ATTTGTGGGC	1470
	GTGTTGCCCTG	CTCCTGGGTT	GGGAAGGGTG	CAGGCCCCAT	GTACCTTCTCT	GTTACTGCCT	TCCAGGTTGG	1540
25	TTCTCAGGGT	TGAATCGTAC	TCGATGTGGT	TTTAGCCCAAC	GGCCCTGCGG	CCAGCTCCTG	GGGGCTGGGG	1610
	AACATGTGTA	AGCACAGAGT	CACCGTCCGC	GTCTTTTGAT	GCCTCACAA	CTCGAGGCTC	CCTGTGTCCG	1680
	TGTTAGTGTG	TGTCACTGTC	CTGCTCACAT	CCTGTCTTGG	GGACGCAGGG	GCTTAGCAGG	TCCCGTAGTA	1750
	AATGACAAGC	GTCTCGGGGG	AGTCTGCAGA	ATAGGAGGTG	GGGGTGCCGG	TCTCTCTCCC	CGCTCTTCAG	1820
	ACTCTTCTCC	TGCCCTGTCT	GTGGCTGCAC	CTGCATCCCT	GCAATCCCTC	CAGCACTGGG	CTGGAGAGGC	1890
30	CCGGGAGCTC	GAGTGCCACT	TGTGCCACGT	GACTGTGGAT	GGCAGTCCGT	CACGGGGGTC	TGATGTGTGG	1960
	TGACTGTGGA	TGGCGGTGTG	TCACAGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2030
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2100
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGTATG	2170
35	TGGTGAAGTG	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGTATG	2240
	TGGTGAAGTG	GGATGGCAGT	CGTGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGTATG	2310
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2380
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2450
	TGTGTGACT	GTGGATGGCG	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	TGGGGTCTGA	2520
40	GGTCTGATGT	GTGGTGAAGT	TGGATGGTGA	TCGGTCAACG	GGGTCTGATG	TGTGTGACT	TCGGGTCTGA	2590
	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	GTGCTGGGGT	TGTGTGACT	TCGGGTCTGA	2660
	GATGGCGATC	GGTCACAGGG	GTCTGATGTG	TGGTGAAGTG	GGATGGCGGT	CGTGGGGTCT	GATGTGTGGT	2730
	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2800
	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2870
45	GACTGTGGAT	GGCGGTCTGTG	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	2940
	TGTGGTGAAG	GGCGGTCTGTG	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	3010
	GTGCTGGGGT	CTGATGTGTG	GTGACTGTGG	ATGGCGGTCTG	GTGCTGGGGT	TGTGTGACT	TCGGGTCTGA	3080
	ATGTGTGGTG	ACTGTGGATG	GCGGTCTGTG	GCTGTGATGT	GCTGTGATGT	GATGGCGGTC	GTGGGTCTGT	3150
	ATGTGTGGTG	ACTGTGGATG	GTGATCGGTC	ACAGGGGTCT	GATGTGTGGT	GACTGTGGAT	GGCGGTCTGT	3220
50	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3290
	GGGTCTGATG	TGTGGTGAAG	GTGGATGGCG	GTGCTGGGGT	CTGATGTGGT	GACTGTGGAT	GGCGGTCTGT	3360
	GCGGTCTGTG	GGTCTGATGT	GTGGTGAAGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	3430
	GCGGTCTGTG	GGTCTGATGT	GTGGTGAAGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	3500
	TGGCGGTCTG	GGGTCTGATG	GTGGTGAAGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	GTGGGTCTGT	3570
55	GCAGGTGGAG	TCCAGGTGTG	GTCTGTAGCT	ACTTTGCGTC	CTCGGCCCCC	CGGCCCCCTG	TTCCCAAAAC	3640
	GAAGCTTCCC	AGGCGCTCTC	TGGGCTTCAT	CCCGCCATCG	GGCTTGGCCG	CAGGTCCACA	CGTCTGTATC	3710
	GGAGAAACAA	AGTGCCCAAG	CTTGGCCGGG	GCAGGCCACA	TTTGTGGCTC	ATGCCCTCTC	CTCTGCGCGG	3780
	AGGTCTCTAC	CTTGACAGAG	CTCCAGCCGT	ACATGCGACA	GTTCTGGGCT	CACCTGCAGG	AGACACAGCC	3850
	GCTGAGGGAT	GCCGCTCGTA	TCGAGCAGGT	CTGGGCACATG	CCCTGCAGGG	TTGGGCACGG	ACTCCACAGA	3920
60	GTGGGTCTCT	CCCTGGGCAA	TCACTGGGCT	CATGACCGGA	CAGACTGTGTG	GCCCTGGGGG	GCAGTGGGGG	3990
	GAATGAGCTG	TGATGGGGGG	ATGATGAGCT	GTGTGCCTTG	GCGAAATCTG	AGCTGGGCCA	TGCCAGGCTG	4060
	CGACAGCTGC	TGCATTACAG	CACCTGTCTCA	CGTTTGACTG	CGCGGCTCTC	CTCCAGTTCC	CGAGTGCCTT	4130
	TGTTTATGAT	TGTGTAATG	TCTTCTCTGC	CAGTTTTGAT	CTTGAGGCCA	AAGGAAAGGT	GTCCCCCTCC	4200
	TTTAGGAGGG	CAGGCCATGT	TTGAGCCGTG	TCCTGCCAG	CTGGCCCTTC	AGTGTGGGT	CTGAGGCCAA	4270
65	AGGAAACGTG	TCCCTTCTCT	TAGGAGGACG	GGCGGTGTTT	GAGCCACGCC	CCGCTGAGCG	GGCTCTCTAG	4340
	TGCTGGGTCT	GTCCACGTGG	CCCTGTGGCC	CTTTGCAGAT	GTGGTCTGTC	CACGTGGGCC	TGTGGCTCTT	4410
	TGCAGATGCC	TGTTAGCACT	TGCTCGGCTC	TAGGGGACAG	TCGTTGCTCC	CGCATGAGGC	TCAGAGACCT	4480
	CTGGGCGAAT	TTCTTGTGCT	CCCAGGGTGG	GGGTGGAGGT	GGCTTGGGCT	GCTGGGACCC	AGACCTCTGT	4550
	CTGGGCGAGT	GGGAGCAAC	TCCTGGATCA	CATATGCCAT	CCGGGCCACG	GTGGGTGTGT	TGGGTGTGAG	4620
70	CCAGCTTGA	CCACAGAGTG	GGCCAGAGGA	GACGTTCTGT	GTACACACAT	CTGCCTAAGC	CCATGTGTGT	4690
	CTGCAGAGAC	TCGGCCCGGC	CAGCCCAAGA	TGGCCCTGCA	TTCCAGCCCA	GCCCGGCACT	TCATCACAAA	4760
	CACCTGACCCC	AAAAGGGAGC	GAGGGTCTTG	GCCACGTGGT	CCTGCTGTGC	TCAGCACCCA	CCGGCTCACT	4830
	CCCATGTGTC	TCCCGTCTGC	TTTCGAGAG	CTCCTCCCTG	AATGAGGCCA	GCAGTGGGCT	CTTCGACGTC	4900
	TTCTACGCT	TCATGTGCCA	CCACGCGGTG	GCATCAGGG	GCAAGTGAGT	CAGGTGGCCA	GGTGCCATTG	4970
	CCCTGCGGGT	GGCTGGGCGG	GCTGGCAGGG	CTTCTGCTCA	CCTCTCTCT	GGCCCTTCCC	CACTGNCCTT	5040

	CTGCCCCGGG	CCACCAGAGT	CTCCTTTTCT	GGCCCCCGCC	CCCTCCGGCT	CCTGGGCTGC	AGGCTCCCGA	5110
	GGCCCCGGAA	ACATGGGCTG	GCTTGGCGCA	GGCGGAGCGG	AGCAGGTGCC	ACACGAGGCC	TGGAAATGGC	5180
	AAGCGGGGTG	TGGAGTTGCT	CCTGCGTGGA	GGACGAGGGG	CGGGGGGTGT	GTCTGGGTCA	GGTGTGCCCC	5250
5	GAGCGTTTGA	GCCTGCAGCT	TGTGAGCTCC	AAGTTACTAC	TGACGCTGGA	CACCCGGCTC	TCACACGGCT	5320
	GTATCTCTCT	CTCCCGATAC	AAAAGGATTT	TATCCGATTC	TCATTCTCTG	CCCTGTCTGT	TGACCCCGCC	5390
	GAGGGCGCGG	GCTCTTCTCT	CTGTGACTAG	ATTTCCCATC	TGGAAAGTGC	GGGGTTGACC	GTGTAGTTTG	5460
	CTCCTCTCGG	GGGGCCTGTG	GTGGCCATGG	GGCAGGCGGC	CTGGGAGAGC	TGCCGTGACA	CAGCCACTGG	5530
	GTGAGGCACA	CTCAGCGTGG	TAGAGCCACA	GTGCTTGGTG	CCACATCAGC	TCCTCTGGAT	TTTAAGTAAA	5600
10	ACCACACACC	TCCCGGCAGG	CATCTGCCTG	CGACCTGTG	TGTGCCCTGG	GAGAGTGGTA	GCACGGAGGA	5670
	AATTCGTGCA	CACCTCAAGG	CATCAGCAAG	GTCTCCGCA	GTCAAGTGGA	ACGTGGAGGC	CTCTCTCTGG	5740
	GATCGCTCC	AGCGGATAAA	GGACTGTGCA	CAGCTTCGGA	AGCTTTTATT	TAAAAATATA	ACTATTAATT	5810
	ATTGCATTAT	AAGTAATCAC	TAATGGTATC	AGCAATTATA	ATATTTATTA	AAGTATAATT	AGAAATATTA	5880
	AGTAGTACAC	ACGTTCTGGA	AAAAACACAA	TTGCACATGG	CAGCAGAGTG	AATTTTGGCC	GAGGGACACG	5950
15	TGTGCACATG	TGTGTAAGCG	GGCCCCAGGC	CCACAGATT	CGCTGACAAA	GTACCTCTCC	CAGAGAAGCC	6020
	ACCACGGGCC	TCCTTCGTGG	TCGTGAATTT	TATTAAGATG	GATCAAGTCA	CGTACCCTCC	ACGTGTGGCA	6090
	GGGCTTTGGG	GAATGTGAGG	TGATGACTGC	CTCCTCATGC	CCTGACAGAC	AGGAGGTGAC	GTGTCTCTGC	6160
	CTGTCCCTAG	GACACGGACA	GGCCCGAAGC	TCTAGTCCCC	ATCGTGGTCC	AGTTTGGCCT	CTGAATAAAA	6230
	ACGTCCTCAA	AACCTGTTGC	CCCAAAACT	AAGAACAGAG	AGAGTTTCCC	ATCCCATGTG	CTCACAGGGG	6300
20	CGTATCTGCT	TGCGTTGACT	CGCTGGGCTG	CCGGGACTCC	TAGAGTTGGT	CGGTGTGCTT	CTGTGCAAAA	6370
	ATGTCAGTCC	TCTTGCCCAT	CACCTGTGTA	TCTGCACAG	CAAGGAAAAG	CTCTTTTCTT	TTCTTTCTTT	6440
	TTTTTTTTTT	GAGACGGAAC	GTGACTGTGG	TCTGCTGGG	CTTGAGTGCA	GTGGCGCGAT	CTCAACTCAC	6510
	TGCAACCTCC	GCCTCCCGGG	TTCCAGCATT	TCTCCTCCCT	CAGCCTCCCG	AGCAGCTGAG	ATTACAGGCA	6580
	CCACCCCTCC	GGCCTGGGCT	AATTTTGTGA	TTTTTAGTAG	AGAGGGGTTT	TGCCCATGTT	GCCAGGGCTG	6650
25	GTCTCGAACT	CCTGACCTCA	GGTGATCCAC	CCACCTCGGC	CTCCCAAAGT	GCTGGGATTA	CAGGTGTGAG	6720
	CCATCAGGCC	CAGCCGGAAA	GCCTCTTTT	AAGGTGACCA	CCTATAGCGC	TTCCCGAAAA	TAACAGGTCT	6790
	TGTTTTTGCA	GTAGGCTGCA	AGCGTCTCTT	AGCAACAGGA	GTGGCGTCTC	GTGGGCTCTG	GTGGGCTCTG	6860
	AGGGTCGCGT	GGCAGCCATG	CCTTCTGTGT	GCACCTTTAG	GTTCACCGGG	GCTATTCTGC	TCTCACTGTT	6930
	TGCTCGAAAA	CGCACCCCTG	GATCCCTTGT	TGGAGAGATT	TCTGCTTCTC	GTGGGTGATG	GTGAACTAGT	7000
30	GGGCAAGGTT	GTATCCGTTG	GGCGCAGCGG	GCTACATGTA	GGGTGATGAG	TCTTTCACCG	TGGACAAATT	7070
	CCTTGAAAAA	AAAAAAGGA	GTCCGGTTAA	GCATTCAATC	CGGGTCAAGT	GTCTGGTTCT	GTGAATAAAT	7140
	TCTAAGATT	AAGAAACCTT	AATGAAAGAA	AACCTTGATG	ATTGAGAGCA	AGGATGTGGT	CACACCTGTG	7210
	GCTGGATCTG	TTTCAGCCGC	CCAGTGATAT	GGTGAGAGTG	GGGAGCAGGG	ATTGTTTGT	CAGAGGTCTC	7280
	ATCTGGTATG	TTTCTGAGGT	GTTTGCCGGG	TGAATGGTAG	ACGTGTCTGT	TGTGTGTATG	AGGTTCTGTG	7350
35	TCTGTGTGTG	GCTCGGTTTG	AGTGTACGCA	TGTCCAGCAC	ATGCCCTGCC	CGTCTCTCAC	CTGTGCTTTC	7420
	CGCGCCACAG	TCCTACGTTT	AGTCCAGGG	GATCCCGCAG	GGCTCCATCC	TCTCCAGGCT	GCTCTGCAGC	7490
	CTGTGCTACG	GGCAGATGGA	GAACAAGCTG	TTTCCGGGGA	TTCGGCGGGA	CGGGTGAGGC	CTCCTCTTCC	7560
	CCAGGGGGGG	TTGGGTGGGG	GTGATTTTGC	TTTTGATGCA	TTCAGTGTTA	ATATTCTTGG	TGCTCTGGAG	7630
	ACCATGACTG	CTCTGTCTTG	AGGAACCGGA	CAAGGTTGCA	GGCCCTTCTT	GGTATGAAGC	CGCAGGGGAG	7700
40	GGGTGTCACA	GCCTGAGGAC	TGGGGGCTCC	ACGCAGGGCTC	TGTCCAGCGG	CCATGTCAGC	AGGCCTCAGG	7770
	GCTCAGCAGG	CGGGAGGGCC	GCTGCCCTGC	ATGATGAGCA	TGTGAATTCA	ACACCGAGGA	AGCACACCA	7840
	CTTCTGTGTC	GTACCCACAG	TTCCGTTAGG	GTCTTGGGG	AGATGGGGCT	GGTGAGGCTT	GAGGCCACCA	7910
	ATCTCCACGC	AGGCCCTCGA	CAGGTGGGCT	GGACTGGGCG	CCTCTTCAGC	CCATTGCCCCA	TCCACTTTCG	7980
	ATGGGGTCTA	CACCCAAGGA	CGCACACACC	TAAATATCGT	GCCAACTTAA	TGTGGTTCAA	CTCAGCTGGC	8050
45	TTTTATTGAC	AGCAGTTACT	TTTTTTTTTT	TAATACTTTA	AGTTCTAGGG	TACATGTGCA	CGACGTGAGC	8120
	GTTAGTTACA	TATGTATACA	TGTGCCATGT	TGGTGTGCTG	CACCCATTAA	CTCATCATTT	ACATTAGGTA	8190
	TATCTCTCAA	TGCTATCCCT	CCCACTCTCC	CCCATCCCAT	GACAGGCCCT	GGTGTGTGAT	GTTCCTCCAC	8260
	CTGTGTCCAA	GTGTTCTCAT	GTTCAGTTT	CCACCTGTGA	GTGAGAACAT	TGTGGTGTGG	GTTCCTTTTC	8330
	CTTGCAATAG	TTTGCTCAGA	GTGATGGTTT	CCAGCTTCGT	CCATGTCCCT	ACAAAGGACA	TGAATCTATC	8400
50	CTTTTTATG	ACTGCATAGT	ATTCGGTGGT	GTATATGTGC	CACATTTTCT	TAATCCAGTC	TATCATCGAT	8470
	GGACATTGGG	GTGGTTTGCA	AGTCTTTGCT	ACTGTGAATA	GTGCCGCAAT	AAACATACGT	GTGCATGTGT	8540
	CTTTATAGCA	GCATGATTTA	TAATCCTTTG	GGTATATACC	CAGTAATGGG	ATGGCTGGGT	CAATGGGTAT	8610
	TTCTAGTTCT	AGATCCTTGA	GGAAATCACCA	CACGTCTTTC	CACAAATGGT	GAACTAGTTT	ACACTCCAC	8680
	CAACAGTGTA	AAAGTGTCTT	GGTGCTGGAG	AGGATGTGGA	CAGCAGTTAT	TTTTTATGTA	AAATAGTATC	8750
55	ACTGAACAAG	CAGACAGTTA	GTGAAGGATG	CGTCAGGAAG	CCTGCAGGCC	ACACAGCCAT	TTCTCTCGAA	8820
	GACTCCGGGT	TTTTCTGTG	CATCTTTTGA	AACCTTAGCT	CCAATTATAG	CATGTACAGT	GGATCAAGGT	8890
	TCTTCTTCAT	TAAGGTTCAA	GTCTAGATT	GAAATAAGTT	TATGTAACAG	AAACAAAAAT	TTCTGTGACA	8960
	CACAACTTGC	TCTGGGATTT	GGAGGAAAGT	GTCTCGAGC	TGGCGGCACA	CTGGTCAGCC	CTCTGGGACA	9030
	GGATACCTCT	GGCCCATGGT	CATGGGGCGC	TGGGCTTGGG	CCTGAGGGTC	ACACAGTGCA	CCATGCCACG	9100
60	CTTCTGTGG	ATAGGATCTG	GGTCTCGGAT	CATGCTGAGG	ACCACAGCTG	CCATGCTGGT	AAAGGGCACC	9170
	ACGTGGCTCA	GAGGGGGCGA	GGTCCCAGC	CCCAGCTTTC	TTACCGTCTT	CAGTTATTTT	TCCCTAAGAG	9240
	TCTGAGAAGT	GGGGCCGCGC	CTGATGGCCT	TCGTTCTGCT	TCAGCTGGCA	CAGAATTGCA	CAAGCTGATG	9310
	GTAACACTG	AGTACTTATA	ATGAATGAGG	AATTGCTGTA	GCAGTTAACT	GTAGAGAGCT	CGTCTGTTGG	9380
	AAAGAAATTT	AAGTTTTTCA	TTTAAACCGCT	TTGGAGAATG	TTACTTTATT	TATGGCTGTG	TAAATGTTT	9450
65	GACATTCAGT	CCCTCGTAGA	CAGATACTAC	GTAATAAGTG	TAAAGTTAAC	CTTGCTGTGT	ATTTTCCCTT	9520
	ATTTTAGGCT	GCTCCTGCGT	TTGGTGGATG	ATTCTTGT	GGTGACACCT	CACCTCACCC	ACGCGAAAAC	9590
	CTTCCCTCAG	TGAGGCCCGT	GCCGTGTGTC	TGTGGGGACC	TCCACAGCCT	GTGGGCTTGG	CAGTTGAGCC	9660
	CCCCGTGTCC	TGCCCTGGC	ACCGCAGCGT	TGCTCTGCCC	AAGTCTCTC	TCTCTGCCGG	TGCTGATACC	9730
	GCAAGAGCAG	AGGCGCTTGG	CCGTGCACCC	AGGCCTGGGG	GGCGAGGGGC	ACCTTCGGGA	GGGAGTGGGT	9800
70	ACCGTGACAG	CCCTGTCTCT	GCAGAGACCG	ACCCAGGTTA	CACACGTGGT	GAGTGCAGGC	GGTGACCTGG	9870
	CTCCTGCTGC	TCTTTGAAAA	GTCAAGAGTG	GCGGCTCCTG	GGGCCCCAGT	GAGACCCCCA	GGAGCTGTGC	9940
	ACAGGGCCTG	CAGGGCCGAG	GCGGCAGCCT	CCTCCCCAGG	GTGACCTGA	GCCTGCGGAG	AGCAGGAGCT	10010
	GCTGAGTGAG	CTGGCCACCA	GCGTTCGCTG	CGGTACAGTT	CCTGCGTGGG	GTTGTTTGGG	ATCGGTGGGA	10080
	GAATTTGGAT	TTGCTGAGTG	CTGCTGTCTT	GAACACGGA	GATGGCTAGG	AGTGGGTTTC	AGAGTTGATT	10150
75	TTTGGAATC	AAACTAAAA	CAGGCACAGG	GGACCTGGCC	TCAGCACAGG	GGATTGTCCA	ATGTGTTCCC	10220
	CCTCAAGGGC	GGCCACACAG	CGCGGTGGGC	TTGTTTTAAA	GTGCGATTGG	ACGAGGGAGC	AGAAACCTTG	10290
	AAAGCTGTAA	AGGGAACCTT	CAGAAATATG	GGCCGCCAGG	GGTGGTTTCA	GGTGCTTTGC	TGGGCTGTGT	10360
	TTGTGAAAC	CCATTTGGAC	CCGCCCTCCA	AGTCACACCT	CCAGGTCCAC	CCTCCAGGGC	CGCCCTGGGC	10430

	TGGGGGTATG	CCTGGCGTTC	CTTGTGCCGC	AGCCCGGAGC	ACAGCAGGCT	GTGCACATTT	AAATCCACTA	10500
	AGATTCACTC	GGGGGGAGCC	CAGGTCCCAA	GCAACTGAGG	GCTCAGGAGT	CCTGAGGGCTG	CTGAGGGGAC	10570
	AGAGCAGACG	GGGAACGCTG	CTTCTGTGTG	GCAAGTTCCT	GAGGGTGCCTG	GCCAGGGAGG	TGGCTCAGAG	10640
	TGTATGTGTT	GGTCCCACCG	GGGGCAGAAC	TCTGTCTCTG	ATGAGTCCGGC	AGCCATGTAA	CAGGAAGGGG	10710
5	TGGCCACAGG	GAGCTGGGAA	TGCACCAGGG	GAGCTGCACA	GCTGGCCGAG	GTCCCAGGGC	CAGGCCACAG	10780
	GAAGGGCAGG	GGGACGCCCC	GGGCCACAGC	AGAGGCCGCA	GGAAGGGAAG	GGGATGCCCA	GGCCAGAGCA	10850
	GAGGCTACCG	GGCACAGGGG	GGCTCCCTGA	GCTGGGTGAG	CGAGGCTCAT	GACTCGGCCA	GGGAACCTCC	10920
	TTGACGTGAA	GCTGACGACT	GGTGTGGCCC	AGCTCACAGC	CCAGCCAGGT	CCCGCGCCTG	AGCAGGAACT	10990
	CAGAACCCCTC	CCCTTTGTCT	AAAGCACAGC	AGATGCCCTC	AGGGCATCTA	GGAGAAAAA	GGCAAACTCG	11060
10	TTGAGAAACG	TCTTAAAGA	AGGTGGGATG	GTGGCAATTT	CTTGTCCAGA	TTTTAGTCTG	CCCCGGACCA	11130
	CAGATGAGTC	TATAACGGGA	TTGTGGTGTT	GCCATGGGGA	CACATGAGAT	GGACCATCAC	AGAGGCCACT	11200
	GGGGCTGCAC	CTCCCATCTG	AGTCTGGCT	GTCCCGGGTC	CAGGCCAGGT	TCTTGCATGC	TCACCTACCT	11270
	GTCTTGCCCG	GGAGACAGGG	AAAGCACCCC	GAAGTCTGGA	GCAGGGCTGG	GTCCAGGCTC	CTCAGAGCTC	11340
	CTGGCAGGCC	CAGCACCCCTG	CTCCAAATCA	CCACTTCTCT	GGGGTTTTCC	AAAGCATTTA	ACAAGGGTGT	11410
15	CAGGTTACCT	CCTGGGTGAC	GGCCCCGCT	CCTGGGGCTG	ACATTGCCCTC	TCTGCCTTAG	GACCTTGGTC	11480
	CGAGGTGTCC	CTGAGTATGG	CTGCGTGGTG	AACTTGGCGA	AGACAGTGGT	GAACTTCCCT	GTAGAAAGCG	11550
	AGGCCCTGGG	TGGCACGGCT	TTTGTTCAGA	TGCCGGCCCA	CGGCCTATTC	CCCTGGTGCG	GCCTGTGCTT	11620
	GGATACCCGG	ACCCTGGAGG	TGCAGAGCGA	CTACTCCAGG	TGAGCGCACC	TGGCCGGAAG	TGGAGCCTGT	11690
	GGCCGGCTGG	GGCAGGTGCT	GCTGCAGGGC	CGTGTGCTCC	ACCTCTGCTT	CCGTGTGGGG	CAGCGGACTG	11760
20	CCAATCCCAA	AGGGTCAGAG	GCCACAGGGT	GCCCTCTGTC	CCATCTGGGG	CTGAGCAGAA	ATGCATCTTT	11830
	CTGTGGGAGT	GAGGGTGCTC	ACAACGGGAG	CAGTTTTCTG	TGCTATTTTG	GTAAGAGGAA	ATGGTGACCC	11900
	AGACCTGGGT	GCATCAGGTT	GTCTTCAGAA	AGCAGTCTGG	ATCCGAACCC	AAGACGCCCG	GGCCCTGCTG	11970
	GGCGTGAGTC	TCTCAAACCC	GAACACAGGG	GCCCTGTGG	GCATGAGTCC	CTCTGAACCC	GAGACCTCTG	12040
25	GGCCCTGCTG	GGCGTGAGTC	TCTCCGAACC	CAGAGACTTC	AGGGCCCTTT	TGGCGCTGAG	TCTCTCCGCT	12110
	GTGAGCCCCA	CACCTCAAAGG	CTCATCCACA	GTCTACAGGA	TGCCATGAGT	TCATGATCAC	GTGTGACCCA	12180
	TGAGGGGACA	GGGCCATGGT	GTGGGGGGGG	TCTCTACAAA	ATTCTGGGGT	CTTGTTTCCC	CAGAGCCCCA	12250
	GAGCTCAAGG	CCCCGTCTCA	GGCTCAGACA	CAATGAATT	GAAGATGGAC	ACAGATGCAG	AAATCTGTGC	12320
	TGTTTCTTTT	ATGAATAAAA	AGTATCAACA	TTCACGGCAG	GGCAAGGTGG	CTCACACCTA	TAATCCCCAG	12390
30	ACTTTGGGAG	CGCGAGGTGG	GTGGATCACT	TGAGGCCAGG	AGTTTGAGGC	CAACCTAAAC	AACATAGTGA	12460
	AATTCCATTT	CTACTTAAAA	AATACAAAA	TTAGCCTGGC	CTGGTGGCAC	ACGCCCTGTAG	TCCCCGCTAT	12530
	CGGGGAGGCT	GAGGCAGGAG	AATCATTGGA	ACCCAGGAGG	CAGAGGTGTC	AGTGAGCCGA	GATCACACCA	12600
	CTGCACTCCA	GCCTGGGCAA	CAGAGTGAGA	CTTCATCTTA	AAAAAAGTATC	AGCATTTCCAA	12670	
	AACCATAGTG	GACAGGTGTT	TTTTTATTCT	GTCCCTCGAT	AATATTTACT	GGTGCTGTGC	TAGAGGGCGG	12740
35	AACTGGGGGT	GCCTTCTCTC	GAAAGGCACA	CCTTCATGGG	AAGAGAAATA	AGTGGTGAAT	GGTTGTAAAA	12810
	CCAGAGGTTT	AACTGGGGT	CCTGTCGTTT	TGAGTTAAAC	GTCCAGATCT	GGACTTTGCC	TCTTTCCAGA	12880
	ATGCTCCCTG	GGGTTTGCTT	CATGGGGGAG	CAGCAGGTGT	GGACACCCCT	GTGATGGGGG	AGCAGCAGGT	12950
	GCAGACGCCC	TCATGATGGG	GGAGTGGCAG	GTGCAGACAC	CCTTGTGCAT	GGTGCCCAAG	ATGTCCCTGT	13020
	TGCAGCTCCC	TCCCCACAAG	GATGCCGGTC	TCTGTGCTCT	CCACAGTCCC	CTGCTTCCCT	CTCACAGCCT	13090
40	TACCTGGTCC	TGGCCTCCAC	TGGCTTTTGT	TGCATGATTT	CCACATTTCC	TGGGCTCCCA	GCACCTCTTC	13160
	GGCTCTCCCA	GGCACCTCTG	CAGTGTGGGC	CATACCAGTC	AGCTGTGAAC	TGTCCACTGC	TTATTTTGCT	13230
	CCCATGAAA	TGTATTTTTT	AGGACAGGCA	CCCTGGTTTC	CAGCCTCTGG	CACAGCATCA	GTGAATGTTA	13300
	TTGAAGGACA	AAGGACAGAC	AAACAATCA	GGAAATGGG	TTCTCTCTAA	ACACATTGCA	AAGCCACAGA	13370
	GGCTAGTGCA	GGATGGGTGG	GCATCAGGTC	ATCAGATGTG	GGTCCAATGC	CAGAATATTC	TGTGCTCCCA	13440
45	AAGGCCACTT	GGTCAGAGTG	TGTGCTTGCA	GAGGTGGCTC	TAAAGGCTCA	GCAGTGGAGG	CAGTGGTTTC	13510
	CCATACTCAG	GGTGAATCTA	CATCCTCTGT	GTCTGAAGTA	TACAGCAGAG	GCTTGAAGGG	CATCTGGGAG	13580
	AAGAAAACAG	GCAAAATGAT	TAAGAAAAGT	GA AAAAGGAA	AAGTGGTAAG	ATGGGAATTT	TCTTGTCCAG	13650
	ATTTTAGTCT	CCCAAAACAC	AGCTCAGATG	GTGAATGTG	GTGAGAACTG	ATGGACAGAA	CAATAGAACA	13720
	AAACGGAAGC	CCTATCTCTC	AGAAACGTGT	GTAAATGTGG	TATGTGGCAC	AGCTGATGGA	AAAGAGAGTG	13790
50	TGTGTGTAAT	TTTTTTTCTT	GAGAAAAGTC	ACTGGAAAGCA	AATAAGTTGT	GTCTTTACAG	CATATACACG	13860
	AGCAGATTCT	AGGTAGAAGA	GGAGACACAT	GCAAAACAAC	CCAGCAACAG	AAATAAAACA	AAAGACTCAA	13930
	AGGGAAGGGA	GGTGAACGTT	CCCTGGTTTG	GTGTTGGGGA	AGGACACACA	GGGAGGGCGA	TGAACCCAGT	14000
	GAGGCAACGG	GCATTGCTTT	CAGTCAGAGT	AAACTCAGCT	TGCCTGAGCC	ACAGTGAATA	TGGCCATTCC	14070
	CTGGAGCGTT	TGTGCACGTG	ATTTATTTAA	GGCGCCCTGT	GAGGTCTCTG	ACATTCACTC	TCTCACTTTG	14140
55	TTCTCCTAAC	CACCTGAGAG	GTAGAGGAGG	AAAGGCTCCA	GGGGAGCAGC	CGCCCTTGGT	CACCCAGCTG	14210
	GCAAAAGGGA	TGCATGATTG	CAGCCTGGCC	TCTGTCTCCG	GGGCGCTTGC	TCTGCCCGAG	GACCCACAC	14280
	AAGTGAGACC	CATAGGCTCA	GGGTGAGCCG	GAGCCCAAGG	TCTGTGTTGG	GATGGCTGTG	AAAGAAAGAA	14350
	TGGACGTCTG	ATGCACACTT	GGGAAGGTCC	TACCAGCAGC	GTCAAAGAAA	TGCATGTGAA	ACTGACAGCG	14420
	AGACCCATCC	CTCAAAGAAA	CGCACGTGAA	ACTGATGGCG	AGACCTGTCC	CCATCCCTCA	TGCTGGCTCC	14490
60	TTTTCTGGGC	TTGCCAAGAG	CCAGCATCAG	GTGAGGCAA	GCTGGAAGA	CTTTCTGGA	AAGCAGCTTG	14560
	TTTGATGGA	AGTCCTCACA	ATGTCCTGTG	TCTTCCAGT	AATTCACACT	CTGAAGTGAC	CAGACATTAT	14630
	CACGGGTCTT	ATTTACCATT	TCCAGTGTTC	CAGGCAGGGG	GACTTGCCAC	AGCAAGTCAC	GAACCTGCC	14700
	AAATACAGGG	CTAAGGAGAT	ATTATGCATC	ACAAAACCTG	CTCTGCCATT	AAACATTTT	CAAGAAATTT	14770
	TTGAAGAATG	TTTAATGGCA	CAAAACGTTT	ATTTCAATGT	AGCAGTGTTC	AAAGCTGGAT	GTAAAAGAAC	14840
65	ACACCCAGG	AGCCTGCCGT	GAATGTCATG	TGTGTTTCTC	TTTGACATG	GACATACATG	GGCAGTGAAT	14910
	GGTGGTGAGG	CCCTGGAGGA	CATCGGTGGG	ATGCCCTCAT	CCTGCCCTTC	TGGAGACACC	ATGTGTGCCA	14980
	CGTGCACTCA	CTGGAGCCCT	GTTTAGCTGG	TGCCACCTGG	CTCTTCCATC	CCTGAGATTC	AAACACAGTG	15050
	AGATTCCCCA	CGCCCAACTC	AGTGTCTTCC	CACAAAAAAC	CTGAGTCACA	CCTGTGTTCA	CTCAGGGGAC	15120
	CGCCGGGAGC	CAGGGCTCCA	CAGTTTATTA	TGTGTTTTTG	GCTGAGTTAT	GTGCAGATCT	CATCAGGGCA	15190
70	GATGATGAGT	GCACAAACAC	GGCCGTGCGA	GGTTTGGAAT	CACCTCAACAT	CACTAGCCAG	GTCTGGTGGG	15260
	AGTTTGGTCA	TGCAGAGTCT	GGATGGCATG	TAGCATTTGG	AGTCCATGGA	GTGAGCACCC	AGCCCTCTCG	15330
	GGCTGCAGCG	CATGCCCCAG	GCAGGACAAG	GAAGCGGGAG	GAAGGCAGGA	GGCTCTTTGG	AGCAAGCTTT	15400
	GCAGGAGGGG	GCTGGGTGTC	GGCCAGGCAC	CTGTGCTGTA	CATTCCCCC	TGTGTCTCAG	CTATGCCCGG	15470
	ACCTCCATCA	GAGCCAGTCT	CACCTTCAAC	CGCGGCTTCA	AGGCTGGGAG	GAACATGCGT	CGCAAACTCT	15540
75	TTGGGGTCTT	CGCGGTGAAG	TGTCACAGCC	TGTTTCTGGA	TTTGACAGGT	AGCAGGCTGA	TGGTCAGCAC	15610
	AGAGTTTCAG	GTTCAGGAGG	TGTGTGCCGA	AGTATGTGTG	TGTGTGTGTG	CGCGCGTGCG	TGCAAGGCTG	15680
	ATGGTGACTG	GCTGCACGTA	AGAGTGACCA	TGTACGCATA	TACACGTGAG	CACATACATG	TGTGATGTG	15750
	TGTACATGAA	GGCATGGCAG	TGTGTGCACA	GGTGTGCAAG	GGCACAAAGT	TGTGCACATG	CGAATGCACA	15820

	CCTGACATGC	ATGTGTGTTT	GTGCACAGTC	GTGTGGGCGAT	TCACGTGAGG	TGCATGCGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTT	ACCCCGCTAG	GTCTCAGCA	15960
	CCAGTGCCAC	TCCTTACAGG	ATGAGACGGG	GTCCAGGGCC	TGGTGGGCT	GAGGCTCTGA	AGCTGCAGCG	16030
5	CTGAGGGCAT	TGTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCTGTGG	GCTTCTGTGT	CCACTCCCCC	16100
	TCTCTGTGG	GCATTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CGCGTCCACT	CCCCCTCTCT	16170
	GTGGGCATCT	CGGTCCACCT	CCCTCTCTGT	TGGGCATTGT	CGTCCACTCC	CTCTCTCTGT	TCCTTCTCTG	16240
	CTTGGCCGAG	CCTCGGGGGC	AGGCAGATGA	CACAGAGTCT	TGACTCGCCC	AGGGTGGTTC	GCAGCTGCCG	16310
	GGTGAGGGCC	AGGCCGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTGTCAA	AATGTTCTCT	TTTCTTGTTC	16380
	CATCTGAATG	GATGATAAAG	CAAAAAGTAA	AAACTTAAAA	TCCCAGAGAG	GTTTCTACCG	TTTCTCACTC	16450
10	TTTCTTGGCG	ACTCTAGGTG	AACAGCCTCC	AGACGGTGTG	CACCAACATC	TACAAGATCC	TCCTGTGCA	16520
	GGCGTACAGG	TGAGCCGCCA	CCAAGGGGTG	CAGGCCAGC	CTCCAGGGAG	CCTCCGCGCT	CTGCTCACCT	16590
	CTGACCCGGG	GGTTACCTTT	GGAACTCCTG	GGTTTATAGG	GCAAGGAATG	TCTTACGTTT	TCAGTGGTGC	16660
	TGCTGCCGTG	GCACAGTTCT	GTTCGCGTGG	CTCTGTGCAA	AGCACCTGTT	CTCCATCTCT	GGTAGTGGT	16730
	AGGAGCCGGT	GTGGCCCCAG	GTGTCCCCAC	TGTGCCTGTG	CACTGGCCGT	GGGACGTGAT	GGAGGCCATC	16800
15	CCAGGGCAGC	AGGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTAGC	AGAGGAGGCT	GGGAAGTGTG	16870
	CTGAACAGTA	GATGGGAGAT	CAGATGCCCG	GAGGATTGGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGGTGC	16940
	AGGTGAGGGT	CGCTGGCCCC	ACCCCGGGGA	AGGTGCAGCA	GAGCTGTGGC	TCCCACACAC	GCCCCGCCAG	17010
	CACCTGTGCT	CTGGGCATGG	CTGTGCTCCT	GGAACTGTTC	CTGTCTGGC	TGGTCAGGGG	GTCCCGCTGC	17080
	CAAGAATCGA	CAACTTTATC	CAAGAGGGAA	GGGCCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCCTGG	17150
20	AGTCAGGGCA	GGTGGTGGCA	CAAGCCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGCACACAC	GGCCCGGGCC	17220
	TCCACCTCAA	CAGGCCCTCC	GAGCCACTGG	GAGCTGAATG	CCAGGAGGCC	GAAGCCCTCG	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTTAC	CCAGGGCCGA	GGCTGCGCGA	ATTACCGTGC	ACACTTGATG	17360
	TGAAATGAGG	TCGTCGTCTA	TCGTGGAAAC	CCAGCAAGGG	CTCAGGGGAG	AGTTTTCCAT	TACAAGGTGC	17430
	TACCATGAAA	ATGGTTTTTA	ACCCGAGTGC	TGCGCCCTTC	ATGCTCTGGC	AGGGAGGGCA	GAGCCACAGC	17500
25	TGCAATGTTAC	CGGCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCAG	GSTGCGTCCG	17570
	GCTCAGACCG	CCCTCTCTCT	TGCTTCTCT	CTCTGCTCA	AATCTTCCCT	CGTTTGCTATC	TCCTTGACCG	17640
	GTGCTGGGC	CCTCGTGCAA	CTGCTTGAC	TCCTTCCGG	AAACCTTGG	GGTGTGCTGG	ATACAGTGTG	17710
	CACTGAGGAC	TGGAGGTGTC	TGACACTGTG	GTGACCCCA	GGGTCCAGCT	GGCGTCTTG	GGGCTCTCTT	17780
	GGGCCATGAT	GAGGTCCAGAG	GAGTTTTCCT	AGGTGAAAC	TCCTGGGAAA	CTCCAGGGG	CATGTGACCT	17850
30	GGCACTGCT	CCTCCCATAT	TCAGCTCAGT	CTTGTCTCA	TTTCCCAACC	AGGGTCTCTA	GCTCCGAGGA	17920
	GCTCCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCAACC	CATGTGGGGA	CCCTTGGGTA	17990
	TCGCTTGAT	TGGGTAGCCC	TGAGGAGGCC	GAGATGCGAT	GGGCCACGGG	CCGTTTCCAA	ACACAGAGTC	18060
	AGGCACGTGG	AAGGCCCAGG	AATCCCTTC	CCTCGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
35	TTTACGGCA	GCCAGGCTGC	AGTGGGCGAG	CGCTGTGGTG	TCCACGTGGC	GCTGGGGGGC	GGGTCTGATT	18200
	CAAAATCCGCT	GGGGCTCGGC	CTTCTGGGCC	CGTGTGGGCC	CGGCTTCCAC	ACGGGCTTGG	GGTGGACGCC	18270
	CGGACCTCTA	GCAGGTGGCT	ATTCTCCCT	TTGGAAGAGA	GGCCCTCACC	CATGCTAGGT	GTTCCTCTCC	18340
	TGGCTCAGGA	CGGTGGCCGT	GTGGCAACCC	CGGACCTTA	GGCTTATTTA	TTTGTTTAAA	AACATTCTGG	18410
	GGCTGGCTTC	CGTTGTTGCT	AAATGGGGAA	AAGACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAAC	18480
40	CGGGGTGCTG	GCTTACTGCT	TGTGATCTCA	GGTCACTCCA	GAAGTGCTCT	AGGAAGTCTG	TGAGACCAGG	18550
	TACATGGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTCAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGG	TCAGGCACCTG	GGTGAGATGA	GGTACACGGG	GGGCTCAGGC	AGAGGGTACG	ACCAGGTACA	18690
	CGGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTGCAC	ATGTGCTGTT	TCATGGTAGC	CAGGTCTGTG	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCCAAAG	ATGGAGGCTG	ACAGGGCTGG	CGCGGTGGCT	18830
45	CACACCTGTA	GTCCAGACAC	TTTGGGAGGC	CGAGGCAGAG	GGATCCCTTG	AGCCCAAGAG	TTTAAGACCA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCATCTCT	ATGAAAAATA	AAAACAAAAA	TAGCTGAAC	ATGGTGGTGT	18970
	CGCGCTGTAG	TTCAATACT	TGGGAGGCTG	AAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAACTGCA	19040
	GTGAGCTGAG	ATTGCACAC	TGTACTGCAG	CCTGGGTGAC	AGAGTGAGAG	CCCATCTCAA	CAACAACAAA	19110
	GAAGACTGAC	AAATGCAAGT	TCTTGGAAAG	AAACATTTAG	TAGGAACTTA	ACCTACACAC	AGAAGCCAGG	19180
	TCGGTGTCTC	GGTGTCAAGT	AGATGAGATG	ATGGGTCTCT	ACACCATCAC	CCCAGACCCA	GGGTTTATGC	19250
50	ACCACAGGGG	CGGGTGGCTC	AGAAGGGATG	CCAGGAGCGT	TGATATACGA	TGACATCAAG	GTTGTCTGAC	19320
	GAAGGGCAGG	ATTATGATA	AGTACCTGCT	GGTACACAAG	GAACAATGGA	TAAACTGGAA	ACCTTAGAGG	19390
	CCTTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTACGCC	19460
	TCCACATGCG	TGTTCATACA	GATGGTGAC	AGAAACCGAG	TGTACCTGTG	CACACACAGA	CACGCAGCTA	19530
	CTCGCACACA	CAAGCACACA	CACAGACATG	CATGCATGCA	TCCGTGTGTG	TGCACCTGTG	CCCATGAGGA	19600
55	AACCCATGCA	TGTGCATTCA	TGCACGCACA	CAGGCACCGG	TGGGCCCATG	CCCACACCCA	CGAGCACCGT	19670
	CTGATTAGGA	GGCCTTTTCT	CTGACGCTGT	CCGCCATCCT	CTCAGGTTTC	ACGCATGTGT	GCTGCAGCTC	19740
	CCATTTTCATC	AGCAAGTTTG	GAAGAACCCC	ACATTTTTTC	TGCGCTCAT	CTCTGACACG	GCCTTCTCT	19810
	GCTACTCCAT	CCTGAAAGCC	AAGAAGCCAG	GTATGTGACG	GTGCTGGGCC	TCAGTGGCAG	CAGTGCTGTC	19880
	CTGCTGGTGT	TAGTGTGTCA	GGAGACTGAG	TGAATCTGGG	CTTAGGAAGT	CTTACCCCT	TTTCGCATCA	19950
60	GGAAGTGGTT	TAACCCAAAC	ACTGTGAGGC	TCGTCTGCCC	GGCCTCTCGT	GGGGTGAAGCA	GAGCACCTGA	20020
	TGGAAGGGAC	AGGAGCTGTC	TGGGAGCTGC	CATCTTCTCC	ACCTTGTCT	GCCTGGGGAA	CGCGTGGGGG	20090
	GCCTGGTCTC	TCCTGTTTGC	CCCATGGTGG	GATTTGGGGG	GCCTGGGCTC	TCCTGTTTGC	CCTGTGGTGG	20160
	GATTGGGCTG	TCTCCGTC	ATGGCACTTA	GGGCCCTTGT	GCAAAACCCAG	GCCAAGGGCT	TAGGAGGAGG	20230
65	CCAGGGCCAG	GCTACCCAC	CCCTCTCAGG	AGCAGAGGCC	CGGTATCACC	ACGACAGAGC	CCCGCGCGCT	20300
	CCTCTGCTTC	CCAGTACCCG	TCCTCTGCCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	CTGAAATTCA	AGCCATGTGC	AACCTGCGGT	CCTGAGCTTA	ACAGCTTCTA	CTTTCTGTTC	TTTCTGTGTT	20440
	GTGGAATTT	CACCTGGAGA	AGCCGAAGAA	AACATTTCTG	TCGTGACTCC	TGCGGTGCTT	GGGTGGGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACCGT	CGGGTGTGGG	CAGCTTTCCG	GTGTCTCCTG	GGAGGGGAGC	20580
70	TGGGCTGGGC	CTGTGACTCC	TCAGCTCTG	TTTTCCCCCA	GGGATGTGCG	TGGGGGCCAA	GGGCGCCGCC	20650
	GGCCCTCTGC	CCTCCGAGGC	CGTGCAGTGG	CTGTGCCACC	AAGCATTCCT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTACCTTA	CGTGCCACTC	CTGGGTGAC	TCAGGACAGG	CAAGTGTGGG	TGGAGGCCAG	TGCGGGCCCC	20790
	ACCTGCCCCAG	GGGTCTCCT	TGAAGCCCT	GTGTGGGGCG	AGCAGCTCA	GATGCTGCTG	AAGTGCAGAC	20860
	GGCCCGGGGC	CTGACCCCTG	GGGCTGGAG	CCACGCTGGC	AGCCCTATGT	GATTAAACGC	TGGTGTCCCC	20930
	AGGCCACGGA	GCCTGGCAGG	GTCCCAACT	TCTTGAACCC	CTGCTTCCCA	TCTCAGGGGC	GATGGCTCCC	21000
75	CAGGCTTGGG	AGCCTTCTGA	CCCTGACCT	GTGCTCTCT	ACAGCCTCTT	CCCTGGCTGC	TGCCCTGAGC	21070
	TCCTGGGGTC	CTGAGCAAGT	TCTCTCCCGG	CCCCGCGCT	CCAGCGTCAC	TGGGTGCTCT	GTCTGCTCGC	21140
	CCCGGTGGAG	GGGTGTCTGT	CCCTTCACTG	AGGTTCCAC	CAGCCAGGGC	CACGAGGTGC	AGGCCCTGCC	21210

TGCCCGGCCA CCCACACGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTCTGGA ACGGAGTCTG 21280
 ATTTTGGCCC CGCAGCCCAG ACGCAGCTGA GTCGGAAGCT CCCGGGGACG ACGTGACTG CCTTGGAGGC 21350
 CGCAGCCAAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT GGCCACCCGC CCACAGCCAG 21420
 GCGGAGAGCA GACACCAGCA GCCCTGTCAC GCCGGGCTCT ACGTCCCAGG GAGGGAGGGG CGGCCACAC 21490
 5 CCAGGCCCGC ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT CCGGCTGAAG 21560
 GCTGAGTGTC CCGCTGAGGC CTGAGCGAGT GTCCAGCCAA GGGCTGAGTG TCCAGCACAC CTGCCGTCTT 21630
 CACTTCCCCA CAGGCTGGCG CTCGGCTCCA CCCCAGGGCC AGCTTTCTCT CACCAGGAGC CCGGCTTCCA 21700
 TCCCCACAT AGGAATAGTC CATCCCCAGA TTCGCCATTG TTCACCCCTC GCCCTGCCCT CCTTTGCCCT 21770
 CCACCCCCAC CATCCAGGTG GAGACCCTGA GAAGGACCTT GGGAGCTCTG GGAATTTGGA GTGACCAAAG 21840
 10 GTGTGCCCTG TACACAGGCG AGGACCCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGGAGGT 21910
 GCTGTGGGAG TAAATACTG AATATATGAG TTTTTCAGTT TTGAAAAAAA TCTCATGTTT GAATCCTAAT 21980
 GTGCACTGCA TAGACACCAC TGTATGCAAT TACAGAAGCC TGTGAGTGAA CCGGGTGGTG GTGAGTGGCG 22050
 GCCCATGGCC TGGCTGTGCA TTTACGGGAG TCTATGAGTG AATGGGGTTG TGGTCAGTGC GGGCCCATGG 22120
 CCTGGCTGGG CCTGGGAGGT TTCTGATGCT GTGAGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190
 15 TAGCCCCAC CCTGGAAGAC ATAACAGTAA GTCCAGGCCG GAAGGGCAGC AGGGATGCTG GGGGCCCAGC 22260
 TTGGCGGCG GGGATGATGG AGGGCTTGGC CAGGGTGGCA GGGATGATGG GGGCCCCAGC TGGGGTGGCA 22330
 GGGGTGATGG GGGGGGCTGG TCTGGGTGGC GGGGAAGATG GGGGAAGCCTG GCTGGGCCCC CTCTCCCCCT 22400
 GCCTCCACCC TGCAAGCCGTG GATCCGGATG TGCTTCCCTG GTGCACATCC TCTGGGCCAT CAGCTTTCAT 22470
 GGAAGTGGGG GGCAGGGGCA TGACACCATC CTGTATAAAA TCCAGGATTC CTCTCTCTGA ACGCCCCAAC 22540
 20 TCAGGTTGAA AGTCACATTC CGCCTCTGGC CATCTCTTAA AGAGTAGACC AGGATTCCTA TCTCTGAAG 22610
 GTGGTAGGG TGGGGCAGTG GAGGGTGTGG ACACAGGAGG CTCAGGGTG GGGCTGGTGA TGCTCTCTCA 22680
 TCTCTTATC ATCTCCACGT CTCATCTCTC ATCTCTTAT CATCTCCAG TCTCATCTGT CTCTCTCTTA 22750
 TCTCCAGTC TCATCTGTCA TCCTCTTACC ATCTCCAGT CTCATCTCTT ATCCTCTTAT CTCTAGTCT 22820
 CATCCAGACT TACCTCCAG GGCGGGTGCC AGGCTCCGAG TGGAGCTGGA CATACGTCTT TCCTCAGGCA 22890
 25 GAAGGAACTG GAAGGATTGC AGAGAACAGG AGGGCGGCT CAGAGGGAGC CAGTCTTGGG GTGAAGAAAC 22960
 AGCCCCCTCT CAGAAGTTGG CTTGGGCCAC ACGAAACCGA GGGCCCTGCG TGAGTGGCTC CAGAGCCTTC 23030
 CAGCAGTCC CTGTTGGGCG CTTATGGTAT GGCCGGGTCC TACTGAGTGC ACCTTGGACA GGGCTCTCTG 23100
 TTTGAGTGCA GCCCGGACGT GCCTGGTGTG GGGGTGGGGC CTTATGGCCA CTGGATATGG GCGTATTAT 23170
 TGCTGTGCT TCAGAGAATG TCTGAGTGAC CGAGCCTAAT GTGTATGGTG GGGCCAAAGT CACAGACTGT 23240
 30 GTCGTAAATG CACTCTGGTG CTTGAGGCC CCGTATAGGA GCTGTGAGGA AGGAGGGGCT CTTGGCAGCC 23310
 GGCCTGGGGG CGCCTTTGCC CTGCAAACTG GAAGGGAGCG GCCCGGGCG CCGTGGGGGG ACGACCTCAA 23380
 GTGAGAGGTT GGACAGAACA GGGCGGGGAC TTCCAGGAG CAGAGGCCGC TGCTCAGGCA CACCTGGGTT 23450
 TGAATCACAG ACCAACAGGT CAGGCCATTG TTCAGCTATC CATCTCTAC AAAGCTCCAG ATTCCTGTTT 23520
 35 CTCGGGGTGT TTTTGTGTA AATTTTACTC AGGATTACTT ATATTTTTTG CTAAGATATT AGACCCTTAA 23590
 AAAAGGTATT TGCTTTGATA TGGCTTAACT CACTAAGCAC CTACTTTATT TGCTGTGTTT TATTTATTAT 23660
 TATTATTATT ATTAGAGATG GTGTCTACTC TGTCACCCAG GTTGTAGTG CAGTGGCACA GTCATGGCTC 23730
 GCTGTAGCCG CAAACCCCA GGCCTCAAGTG ATCCTCCGGC CTCAGCTTCC CAGAGTGCTG GGATTACAGG 23800
 TGTGAGCCAC TGCCCTTGCC TGGCACTTTT AAAAACCCT ATGTAAGGTC AGGTCCAGTG GCTTCCACAC 23870
 40 CTGTATCCC AGTAGTTTGG GAAGCCGAGG CAGAAGGATT GTCTGAGGCC AGGAGTTTGA GACCAGCATG 23940
 GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAA AGTTATCCGG GCGTGGGGTC CAGCATCTGT 24010
 AGTCCCAGCT GCTCGGGAGG CTGAGTGGGA GGATCGCTTG AGCCCGGGAG GTCATGGCTG CAGTGAGCTG 24080
 TGATTGTACC ATCGCACTCG AGCCTGGGCA ACAGAGTGAG ACCCTGTCTC AAAAAAAAAA AAAAAAAAAA 24150
 AAGGAGAAGG AGAAGAGAAG AAGAGGAAG AAGGAAAGAG AAGAAGAAG AAGAAGAAG AAGAAGGAG 24220
 45 AAGGAGCCCT GCTAGGTGCT AGGTAGACTG TCAATCTCA GAGCAAAATG AAAATAACAA AGTTTTAAG 24290
 GGAAAGAAAA ACCCCAGCTC TTTGACTTTC CTTAGGCCCTG AACTTCTATC CAAGCAGCTT CCTTCCACAG 24360
 ACAAGCGTGT ATGGAGCGAG TGAGTTCAAA GCAGAAAGGG AGGAGAAGCA GGCAAGGGTG GAGGCTGTGG 24430
 GTGACACCA GAGGAGCCCG TGAAGGGAG TGGTTGTTTT CTTGCCCTCAG CCCACGCTC CTGCCGGTCC 24500
 TGCACTGCT GTAACCGTGC ATGTTGGTGC CAGGTGCCCA CTTGGGAAGG ATGCTGTGCA GGGGGCTTGC 24570
 50 CAAACTTTGG TGGGTTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCTC CCCAAGATG 24640
 CCCACGTCCT TCTCTGGAA CTTGTGAATG TGTACCCGC AAGGCAGAGG CTGGTGAAGG CTGAGGTGG 24710
 AATCACGGCT GCCAGTCAGC CGATCTTAAG GTCATCTGG ATTATCTGT GGGCCTGATA TGGCCACAAG 24780
 GGTCCCTAGA AGTGAGAGAG GGAGGCAGGG GAGAGTCAGA GAGGGGACGT GAGAAGGACC ACTGGCCACT 24850
 GCTGGCTTTG AGATGGAGGA GGGGGTCCCC AGCCAAGGAA TGGGGGCAGC CGCTCCATGC TGGAAAAGCA 24920
 55 AGCAATCCTC CCCGCTCCTG AGGGCACACG GCCCTGCCCA CGCCTCGATT TCAGGCCAGT GGGACCTGTT 24990
 TCAGCTTTCC GGCCTCCAGA GCTGTAAGAT GATGCGTTTG TGTTACGCCA CTAAGCTGCA GTGATTCGTC 25060
 ACAGCAGCAA ATGGAATAGC AGTACAGGGA AATGAATACA GGGACAGTTC TCAGAGTGAC TCTCAGCCCA 25138
 CCCCTGGG

60 Beispiel 5

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz
 der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der
 Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist
 65 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in
 5 den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk.(%)	70	60	80		100	100	95	70			80	100	100		60

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616
 15 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);

Intron 2: bp 12951-21566 (SEQ ID NO 5);

Intron 3: bp 21763-23851 (SEQ ID NO 6);

5 Intron 4: bp 24033-24719 (SEQ ID NO 7);

Intron 5: bp 24900-25393 (SEQ ID NO 8);

5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);

Intron 7: bp 3879-4858 (SEQ ID NO 11);

Intron 8: bp 4945-7429 (SEQ ID NO 12);

Intron 9: bp 7544-9527 (SEQ ID NO 13);

15 Intron 10: bp 9600-11470 (SEQ ID NO 14);

Intron 11: bp 11660-15460 (SEQ ID NO 15);

Intron 12: bp 15588-16467 (SEQ ID NO 16);

Intron 13: bp 16530-19715 (SEQ ID NO 17);

Intron 14: 19841-20621 (SEQ ID NO 18);

20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position
21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCCGGGTCCGGCTCCGGTGGGGTTGAGGGCGGCCGGGGGAACCAGCGACATGCGGAGAGCAGCGCAGG
CGACTCAGGGCGCTTCCCCCGCAG

5

Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCCGTCGAGGGCCAGGCCCCAGAGCTGAATGCAGTAGGGGCTCAGAAAAGGGGGCAGGCAGAGCC
CTGGTCCTCCTGTCTCCATCGTCACGTGGGCACACGTGGCTTTTCGCTCAGGACGTGAGTGACACGGTGATCTCTGCC
TCTGCTCTCCCTCCTGTCCAGTTTGCATAAACTTACGAGGTTACCTTCACGTTTGTATGGACACGCGGTTTCAGGCGC
10 CGAGGCCAGAGCAGTGAACAGAGGAGGCTGGGCGCGGCAGTGGAGCCGGTTGCCGGCAATGGGAGAAAGTGTCTGGAAG
CACAGACGCTCTGGCGAGGGTGCTGCAGGTTACCTATAATCCTCTTCGCAATTTCAAGGTTGGGAATGAGAGGTGGGGA
CGAGAACCCCTCTTCTGGGGTGGGAGGTAAGGGTTTTGCAGGTGCACGTGGTCAGCCAATATGCAGGTTGTGTTTA
AGATTTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTCACGCCGTAATCCAGCACTTTGGGAAGCTGAGGCAGGTGGA
15 TCACCTGAGGTGAGGAGTTTGTAGACAGCCTGACCAACATGGTGAAACCCTATCTGTACTAAAAATACAAAATTAGCTG
GGCATGGTGGTGTGTGCCTGTAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGC
TGCAGTGAGCTGAGATTGTGCCATTGTACTCCAGCCTGGGCGACAAGAGTGAAACTCTGTCTTTAAAAAAGTGTGTT
CGTTGATTGTGCCAGGACAGGGTAGAGGGAGGGAGATAAGACTGTTCTCCAGCACAGATCCTGGTCCCCTCTTTAGGTAT
GAAGAGGGCCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGTGGTGTTCAGGGG
ATGGTGTCTGGGCCCTGCCGTGTCCCCACCTGTTTTCTGGATTGATGTTGAGGAACCTCCGCTCCAGCCCCCTTT
20 TGGCTCCAGTGCTCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCCGATTTCACCCCCCTCCCAAACTCCCAAGAC
ATGTAAGACTTCCGGCCATGCAGACAAGGAGGGTGACCTTCTTGGGGCTCTTTTTTCTTTTTTCTTTTTATGGTGGC
AAAAGTCATATAACATGAGATTGGCACTCCTAACACCGTTTTCTGTGTACAGTGAGAATTGCTAACTCGGCGGTGTTTA
CAGCAGGTTGCTTGAATGCTGCGTCTTGCGTGACTGGAAGTCCCTACCCATCGAACGGCAGCTGCCTCACACCTGCTGC
GGCTCAGGTGGACCACGCCAGTCAAGTACAGTAAAGCGTCATGCAACCCAGTTTTGCTTTTTGTGCTCCAGCTTCCTTCGTTGAG
25 GAGAGTTTGAGTTCTCTGATCAGGACTCTGCCTGTCTATGCTGTTCTCTGACTTCAGATGAGGTACAACTCTGCCCTGG
CTTATGCAGGGAGTGAGGCGTGGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGTTGCCCCAGGTGTCCCT
GTCACGTGTAGGGTGAGTGAGGCGCGGCCCCGGGTGTCCCTGTCCCGTGCAGCGTGATTGAGGTGTGCCCCCGGGTGT
CCCTGTACGTGTAGGGTGAGTGAGGCGCCATCCCCGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCGGTGTCCTCGG
GTGTCCCTGTCCCGTGCAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGGCGCGGTCC
30 CCGGGTGTCCCTCTCAGGTGTAGGGTGAGTGAGGCGCGGCCCCAGGGTGTCCCTGTACGTGTAGGGTGAGTGAGGCACC
GTCCCTGGGTGTCCCTCCAGGTATAGGGTGAGTGAGGCACTGTCCCCGGGTGTCCCTGTACGTGCAGGGTGAGTGAGG
CGCGGCCCCGGGTGTCCCTCTCAGGTGCAGGGTGAGTGAGGCGCTGTCCCTGGGTGTCCCTGTCTCGTGTAGGGTGAGT
GAGGCTCTGTCCCAGGTGTCTTGGCGTTTGCTCACTTGAGCTTGCTCCTGAATGTTTGCTCTTTCTATAGCCACAGCT
GCGCCGGTTGCCATTGCCTGGGTAGATGGTGCAGGCGCAGTGCTGGTCCCCAAGCCTATCTTTTCTGATGCTCGGCTCT
35 TCTTGGTCACCTCTCCGTTCCATTTTGCTACGGGGACACGGGACTGCAGGCTCTCGCCTCCCGGTGCCAGGCACTGCAG
CCACAGCTTCAGGTCCGCTTGCTCTGTTGGGCTGGCTTGTCAACACGTGCCCGCCACATGCTGCTGCAATACTCC
TCTCCAGCTTGTCTCATGCCAGGCTGGACTCTGGGCTGCCTGTGTCTGCTGCCACGTGTTGTGGAGACATCCAGAA
AGGGTTCTGTGTCCCTGAAGGAAAGCAAGTACCCCCAGCCCCCTCACTTGTCTGTTTTCTCCCAAGCTGCCCTCTGCG
TTGGCCCCCTTGGGTGGGTGGCAACGCTTGTACCTTATTCTGGGCACCTGCCGCTCATTGCTTAGGCTGGGCTCTGCTT
40 CCAGTCGCCCCCTCACATGGATTGACGTCCAGCCACAGGTGGAGTGTCTCTGTCTGCTCCTGCTCTGAGACCCACGTG

GAGGGCCGGTGTCTCCGCCAGCCTTCGTGAGACTTCCCTCTTGGGTCTTAGTTTTGAATTTCACTGATTTACCTCTGACG
TTTCTATCTCTCCATTGTATGCTTTTTCTTGGTTTATTCTTTCATTCTTTCTAGCTTCTTAGTTTAGTCATGCCTTTC
CCTCTAAGTGCTGCCTTACCTGCACCCGTGTTTTGATGTGAAGTAATCTCAACATCAGCCACTTTCAAGTGTTCTTAAA
ATACTTCAAAGTGTTAATACTTCTTTTAAGTATTCTTATTCTGTGATTTTTCTTTGTGCACGCTGTGTTTTGACGTGA
5 AATCATTTTGATATCAGTGACTTTTAAGTATTCTTTAGCTTATTCTGTGATTTCTTTGAGCAGTGAGTTATTTGAACACT
GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTTAAGTTATCATTTTATTATTGATTTCTAACTCAGT
TGTGTAGTGGTCTGTATAATACCAATTATTTGAAGTTTGCAGGAGCCTTGCTTTGTGATCTAGTGTGTGCATGGTTTCCAG
AACTGTCCATTGTAAATTTGACATCCTGTCAATAGTGGGCATGCATGTTCACTATATCCAGCTTATTAAGGTCCAGTGCA
AAGCTTCTGTCTCCTTCTAGATGCATGAAATCCAAGAAGGAGGCCATAGTCCCTCACCTGGGGGATGGGTCTGTTCAAT
10 TCTTCTCGTTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGTGGTAGAATTTTATCTTCTGATGAGTGAA
TCTTTTGGAGACTTCTATGTCTCTAGTAATCTAGTAATTTCTTTTTTAAATTGCTCTTAGTACTGCCACACTGGGCTTCT
TTTGATTAGTATTTTCTGTGTGTCTGTTTTCTGCCTTTAATTTATATAATATAATATATTTTTTTTTTTTGGAGACA
GAGTCTTGGTCTGTGCGCCAGGGTGAGTGAGTGGTGTGATCACAGGTGAGTGAACTTTTACCTTCTGGCCTGAGCCGT
CCTCTCACCTCAGCCTCCTGAGTAGCTGGAACGACAGACACGACCGCTACACCTGGCTAATTTTTAAATTTTTCTGGA
15 GACAGGGTCTTGCTGTGTGCGCCAGGCTGGTCTCAAACCTCTTGACTCAAGGGATCCATCTACCTCGGCTTCCCAAAGTG
CTGAATTACAGGCATGAGCCACCATGTCTGGCCTAATTTTCAACACTTTTATATTCTTATAGTGTGGGTATGTCTGTTA
ACAGCATGTAGGTGAATTTCCAATCCAGTCTGACAGTCGTGTTTAACTGGATAACCTGATTATTTTCAATTTTTTGTG
ACTAGAGACCCGCTGGTGCATCTGATTCTCCACTTGCCCTGTTGCATGTCTCGTTCCCTGTTTCTCACCACCTCTTG
GGTGGCATGTGCGTTTCTGCGAGTGTGTGATCCTCTCGTTGCTCCTGGTCACTGGGCATTTGCTTTTATTCT
20 CTTTGCTTAGTGTTACCCCTGATCTTTTTATTGTCGTTGTTGCTTTTGTATTGAGACAGTCTCACTCTGTCAACCA
GGCTGGAGTGTAAATGGACAATCTCGGCTCACTGCAACCTCTGCTCCTCGGTTCAAGCAGTTCTCATTCCTCAACCTCA
TGAGTAGCTGGGATTACAGGCGCCACCACCAGCCTGGCTAATTTTGTATTTTAGTAGAGATAGGCTTTCACCATGT
TGGCCAGGCTGGTCTCAAACCTCTGACCTCAAGTGATCTGCGCCGCTTGGCCTCCACAGTGCTGGGATTACAGGTGCAA
GCCACCGTGCCCGCATACCTTGATCTTTTAAATGAAGTCTGAAACATGTCTACCTTGTCTGAGCAATAAGACCCCT
25 AGTGATTTTAGCTCTGGCCACCCCGAGCCTGTGTGCTGTTTTCCCTGCTGACTTAGTCTATCTCAGGCATCTTGACA
CCCCACAAGCTAAGCATTATTAATATTGTTTTCCGTGTGAGTGTCTGTAGCTTTGCCCCCGCCCTGCTTTCTCTC
TTTGTTCCTGCTGTCTCTGCTCAGGCGCGCGTCTGGGGTCCCTTCTGTCTTTGCGTGGTCTCTGTCTGTG
TTATTGCTGGTAAACCCAGCTTTACCTGTGCTGGCCTCCATGGCATCTAGCGAGCTCCGGGACCTCTGCTTATGATGC
ACAGATGAAGATGTGGAGACTCACGAGGAGGCGGTCTCTTGGCCCGTGTGCTGAGTGTCTGGAGCACCAGTGGCCAGCGTTC
30 CTTAGCCAGTGAGTGACAGCAACGTCGCTCGGCTGGGTTGAGCCTGGAAAACCCAGGCATGTGGGGTCTGGTGGCT
CCGCGGTGTGAGTTTGAATCGCGCAAACCTGCGGTGTGGCGCCAGCTCTGACGGTGTGCTGGCGGGGAGTGTCTG
CTTCTCCTTCTGCTTGGGAACAGGACAAAGGATGAGGCTCCGAGCCGTTGTGCGCCAACAGGAGCATGACGTGAGCC
ATGTGGATAATTTTAAATTTCTAGGCTGGGCGCGGTGGCTCACGCTGTAATCCAGCACTTTGGGAGGCCAAGGCGGG
TGGATCACGAGGTCAGGAGGTCGAGACCATCTGGCCAACATGATGAAACCCATCTGTACTAAAAACACAAAAATTAGC
35 TGGGCGTGGTGGCGGTGCTGTAATCCAGCTACTCGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGTTGGAA
GTTGCAGTGAGCCGACATTGCACCACTGCCTCCAGCCTGGCAACACAGCGAGACTCTGTCTCAAAAAAAAAAAAAA
AAAAAAAAAAATTTCTAGTAGCCATTAATAAAGTAAAAAAGAAAAGGTGAAATTAATGTAATAATAGATTTTACTGAA
GCCCAGCATGTCCACACCTCATATTTAGGGTGTATTGGTGGGAGCATCACTCACAGGACATTTGACATTTTTTGGAGC
TTTGTCTGCGGGATCCCGTGTGTAGGTCCCGTGCCTGGCCATCTCGGCTGGACCTGTGGGCTTCCATGGCCATGGCT
40 GTTGTACCAGATGGTGAGGTCCGGATGAGGTGCGCCAGGCCCTCAGTGAGCTGGATGTGAGTGTCCGGATGGTGACG
TCTGGATGAGGTGCGCCAGGCCCTGCTGTGAGCTGGATGTGTGGTGTCTGGATGGTGACGCTCAGGGGTGAGGTCTCCAG

5 G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G
T G T G G T G T C T G G A T G G T G C A G G T C A G G G G T G A G G T C T C C A G G C C C T C G G T A A G C T G G A G G T A T G G A G T C C G G A T G A T G C A
G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G T G T G T C T G G A T G G T G C A G G T C T G G G G T G A G G T C A C C
A G G C C C T G C G G T G A G C T G G G T G T C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G A C G G T G C C A G A C C A T G C
10 G G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T T G C C A G G C C C T G C T G T G A G T T G G A T G T G G G G T
G T C C G G A T G C T G C A G G T C C G G T G T G A G G T C A C C A G G C C C T G C T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T
G G G G T G A A G G T C G C C A G G C C C C T G C T T G T G A G C T G G A T G T G T G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G
G C C C T C G G T G A G C T G G A T G T G C A G T G T C C A G A T G G T G C A G G T C C G G G T G A G G T C G C C A G A C C C T G C G G T G A G C T G G A T G
T G C G G T G T C T G G A T G G T G C A G G T C T G G A G T G A G G T C G C C A G G C C C T C G G T G A G C T G G A T G T A T G G A G T C C G G A T G G T G C C
15 G G T C C G G G T G A G G T C G C C A G A C C C T G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T A C A G G T C T G G A G T G A G G T C G C C
A G A C C C T G C T G T G A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T C A G G G G T G A G G T C T C C A G G C C C T C G G T G A G C T G G A
G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A A C T G G A T G T G C G G C G T C T G G A T G G T
G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G T G A G G T C G
C C A G G C C C T G C T G T G A G C T G G A T G T G C G G C G T C T G G A T G G T G C A G G T C T G G G G T G T G G T C G C C A G G C C C T C G G T G A G C T G
20 G A G G T A T G G A G T C C G G A T G A T G C A G G T C C G G G G T G A G G T T G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T
G T G C A G T C C G G G T G A G G T C G C C A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G G G T G A G G T
C A C C A G G C C C T G C G G T G A G C T G G T T G T G C G G T G T C C G G T T G C T G C A G G T C C G G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T G T G C G G T G T C C C C G T G T C C G G A T G G T G C A G G T C C A G G G T G A G G T C G C T A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C T T T G G T G A G C T G G A T G T G C G G T G T C T G C A T G G T G C A G G T C T G
25 G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G T G G T G T C C G G A T G G T G C A G G T C C G G C G T G A G G T C G C C A G G C C C T
G C T G T G A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G G T G A G G T A G C C A A G G C C T T C G G T G A G C T G G A T G T G G G
G T G T C C G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C C G G A T G G T G C A G G T
C C G G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C T G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G
C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C A G T G A G C T G G A T G
30 T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T T A G C T G G A T A T G C G G T G T C G G A T G G T G C A
G G T C C G G G T G A G G T C A C C A G G C C C T G C G G T T A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C
A G G C C C T G C T G T G A G C T G G A T G T G C T G T A T C C G G A T G G T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G
A T G T G C T G T A T C C G G A T G G T G C A G G T C T G G C G T G A G G T C G C C A G G C C C T G C G G T G A G C T G G A T G T G C A G T G T A C G G A T G G
T G C A G G T C C G G G G T G A G G T C G C C A G G C C C T G C G G T G G G C T G T A T G T G T G T G T C T G G A T G G T G C A G G T C C G G G G T G A G T T
35 C G C C A G G C C C T G C G G T G A G C T G G A T G T G T G G T G T C T G G A T G C T G C A G G T C C G G G G T G A G T T C G C C A G G C C C T C G G T G A G C
T G G A T A T G C G G T G T C C C C G T G T C C G A A T G G T G C A G G T C C A G G G T G A G G T C G C C A G G C C C T T G G T G G G C T G G A T G T G C C G T
G T C C G G A T G G T G C A G G T C T G G G G T G A G G T C G C C A G G C C C T T G G T G A G C T G G A T G T G C G G T G T C C G G A T G G T G C A G G T C C G
G G G T G A G G T C A C C A G G C C C T C G G T G A T C T G G A T G T G G C A T G T C C T T C T C G T T T A A G

35 Intron 3 (SEQ ID NO 6)

G T A C T G T A T C C C C A C G C C A G G C C T C T G C T T C T G A A G T C C T G G A A C A C C A G C C C G G C C T C A G C A T G C G C C T G T C T C C A C T
T G C C T G T G C T T C C C T G G C T G T G C A G C T C T G G G C T G G G A G C C A G G G C C C C G T C A C A G G C C T G G T C C A A G T G G A T T C T G T G
C A A G G C T C T G A C T G C C T G G A G C T C A G T T C T C T T A C T T G T A A A A T C A G G A G T T T G T G C C A A G T G G T C T C T A G G G T T T G T A
A A G C A G A A G G G A T T T A A A T T A G A T G G A A C A C T A C C A C T A G C C C T C C T T G C C T T T C C C T G G G A T G T G G G T C T G A T T C T C T C
40 T C T C T T T T T T T T C T T T T T G A G A T G G A G T C T C A C T C T G T T G C C C A G G C T G G A G T G C A G T G G C A T A A T C T T G G C T C A C T

5 GCAACCTCCACCTCCTGGGTTTAAAGCGATTACCAGCCTCAGCCTCCTAAGTAGCTGGGATTACAGGCACCTGCCACCAC
GCCTGGCTAATTTTGTACTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTCATGACCTCAGG
TGATCCACCCACCTTGCCCTCCCAAAGTGCTGGGTTTACAGGCTAAGCCACCGTGCCAGCCCCGATTCTCTTTAATT
CATGCTGTTCTGTATGAATCTTCAATCTATTGGATTAGGTCATGAGAGGATAAAATCCACCCACTTGGCGACTCACTG
CAGGGAGCACCTGTGCAGGGAGCACCTGGGGATAGGAGAGTTCCACCATGAGCTAACTTCTAGGTGGCTGCATTGGAATG
GCTGTGAGATTTTGTCTGCAATGTTTCGGCTGATGAGAGTGTGAGATTGTGACAGATTCAAGCTGGATTTCATCAGTGAG
GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTTCTCCGTGTCCGCCCAGGC
TGACTGTGGAGGGCTTTAGTCAGAAGATCAGGGCTTCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGGCTTGTGAC
ACCCCATGCCCAAATCAGGATGTCTGCAGAGGGAGCTGGCAGCAGACCTCGTCAGAGGTAACACAGCCTCTGGGCTGGG
10 GACCCCGACGTGGTGTGGGGCCATTTCTTGTCATCTGGGGAGGGTCAGGGCTTTCCTGTGGGAACAAGTTAATACAC
AATGCACCTTACTTAGACTTTACACGTATTTAATGGTGTGCGACCCAAACATGGTCATTGACCAGTATTTTGAAAGAAT
TTAATTGGGGTGACCGGAAGGAGCAGACAGACGTGGTGGTCCCAAGATGCTCCTTGTCACTACTGGGACTGTTGTTCTG
CCTGGGGGGCTTGGAGGCCCTCTCCCTGGACAGGGTACCGTGCTTTTCTACTCTGCTGGGCTGCGGCTGCGGTC
AGGGCACCAGCTCCCGAGCACCCGCGGCCAGTGTCCACGGAGTGCCAGGCTGTGAGCCACAGATGCCAGGTCCAGGT
15 GTGGCGCTCCAGCCCCGTGCCCCATGGGTGGTTTGGGGGAAAAGGCCAAGGCGAGAGGTGTGAGGAGTGTGGTGGG
CTCATGAGAGCTGATTCTGCTCCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCCGCCTCTCCATCTGAAGGGATGTGGCT
CTTTCTACCTGGGGGTCTGCTGGGGCCAGCCTTGGGTACCCAGTGGCTGTACCAGAGGGACAGGCATCTGTGTGG
AGGGGCATGGGTTACGTGGCCCCAGATGCAGCCTGGGACCAGGCTCCCTGCTGTGATGGTGGGACAGTACCCTGGGG
GTTGACCGCGGACTGGGCGTCCCAGGGTTGACTATAGGACCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAG
20 AGGCGTCTGGCTGGCATGGGTGGACGTGGCCCCGGGCATGGCCTTCAGCGTGTGCTGCCGTGGGTGCCCTGAGCCCTCAC
TGAGTCGGTGGGGCTTGTGGCTTCCCGTGAGCTTCCCCCTAGTCTGTTGTCTGGCTGAGCAAGCCTCCTGAGGGGCTCT
CTATTGCAG

Intron 4 (SEQ ID NO 7)

25 GTGGCTGTGCTTTGGTTTAACTTCCTTTTAAACAGAAAGTGCCTTTGAGCCCCACATTTGGTATCAGCTTAGATGAAGGG
CCCCGAGGAGGGGCCACGGGACACAGCCAGGGCCATGGCACGGCGCCAACCCATTGTGCGCACAGTGAGGTGGCCGAGG
TGCCGGTGCCCTCAGAAAAGCAGCGTGGGGGTGTAGGGGGAGCTCCTGGGGCAGGGACAGGCTCTGAGGACCACAAGAAG
CAGCCGGGCCAGGGCCTGGATGCAGCACGGCCCCAGGTCTGGATCCGTGTCCTGCTGTGGTGCAGCCTCCGTGCGCT
TCCGCTTACGGGGCCCGGGGACCAGGCCAGCTGCCAGGAGCCCCACCGGGCTCTGAGGATCCTGGACCTTGCCCCACGG
30 CTCCTGCACCCACCCCTGTGGCTGCGGTGGCTGCGGTGACCCGTCATCTGAGGAGAGTGTGGGTGAGGTGGACAGAG
GTGTGGCATGAGGATCCCGTGTGCAACACACATGCGGCCAGGAACCGTTTCAAACAGGGTCTGAGGAAGCTGGGAGGGG
TTCTAGGTCCCGGTCTGGGTGGCTGGGGACACTGGGGAGGGGCTGCTTCTCCCTGGGTCCCTATGGTGGGTGGGCAC
TTGGCCGGATCCACTTTCCTGACTGTCTCCCATGCTGTCCCCGCCAG

35 Intron 5 (SEQ ID NO 8)

GTGGGTGCCGGGACCCCCGTGAGCAGCCCTGCTGGACCTTGGGAGTGGCTGCCTGATTGGCACCTCATGTTGGGTGGAG
GAGGTACTCTGGGTGGGCGCAGGGAGTGCAGGTGACCTGTCACTGTTGAGGACACACCTGGCACCTAGGGTGGAGGC
CTTCAGCCTTTCCTGCAGCACATGGGGCCGACTGTGCACCTGACTGCCCGGGCTCCTATTCCCAAGGAGGGTCCCACTG
GATTCAGTTTCCGTGAGAGAAGGAACCGCAACGGCTCAGCCACCAGGCCCGGTGCCCTGCACCCAGTCTCTGAGCCAG
40 GGGTCTCCTGCTCCTGAGGCTCAGAGAGGGGACACAGCCCGCCTGCCCTTGGGGTCTGGAGTGGTGGGGGTGAGAGAG

AGTGGGGGACACCGCCAGGCCAGGCCCTGAGGGCAGAGGTGATGTCTGAGTTTCTGCGTGGCCACTGTCACTCTCCTCGC
CTCCACTCACACAG

5'-Bereich Intron 6 (SEQ ID NO 9)

5 GTAAAGGTTACGTGTGATAGTCGTGTCCAGGATGTGTGTCTCTGGGATATGAATGTGTCTAGAATGCAGTCGTGTCTGTG
ATGCGTTTCTGTGGTGGAGGTACTTCCATGATTTACACATCTGTGATATGCGTGTGTGGCACGTGTGTGCTGGTGCAT
GTATCTGTGGCGTGCATATTTGTGGTGTGTGTGTGTGGCACGTGTGTGTCCATGGTGTGTGCTGTGGTGTGCATG
TGTGTGTGTCTGTGACACGTGCATGTTTCATGCTGTGTGCTGCATGTCTGTGATGTGCCTATTTGTGGTGTGTGTGCAT
GTGTCCGTGACATATGCGTGTCTATGGCATGGGTGTGTGTGGCCCCCTTGGCCTTACTCCTTCTCCTCCAGGCATGGTCC
10 GCACCATTTGTCTCACGCTCTCGGGTGTGGTTTGGGGAGCTCCACATTCAGGGTCTCACTTCTAGCATGGGTGCCCT
GTCCTGTACAGGGCTGGGCCCTTGGAGACTGTAAGCCAGGTTTGAGAGGAGAGTAGGGATGCTGGTGGTACCTTCCTGGA
CCCCCTGGCACCCCCAGGACCCAGTCTGGCCTATGCCGGCTCCATGAGATATAGGAAGGCTGATTAGGCCTCGCTCCCC
GGGACACACTCTCCAGAGCGGCCGGGGCTTGGGGCTCGGCAGGGGTGAAAGGGGCCCTGGGCTTGGGTTCCACCC
AGTGGTCATGACACGCTGGAGGGTAAGCCCTCAAAGTCGTGCCAGGCCGGGTGACAGGTGAAGAAGTATCCCTGGA
15 GCTTCGGTCTGGGGAGAGGCACATGTGGAACCCACAAGGACCTCTTCTCTGACTTCTTGAGCT

3'-Bereich Intron 6 (SEQ ID NO 10)

TGTGGGATTGGTTTTTCATGTGTGGATAGGTGGGATCTGTGGGATTGGTTTTTATGAGTGGGGTAACACAGAGTTCAAG
GCGAGCTTTCTCCTGTAGTGGGTCTGCAGGTGCTCCACAGCTTTATTGAGGAGACCATATCTTCTTTGAACATAGGT
20 CGGGTTTATAGTAAGTCAGGGGTGTGGAGGCTCCCTGGGCTCCCTGTTCTGTTTCTTCCACTCTGGGGTCTGTGGT
CCTGCTGTGGTGTGTGGCCGGTGGGCAGGGCTTCCAGGCTCCTTGTGTTCAATGGCCTGGATGTGGCCCTGGCTACGCT
CCGTCTTGGAAATCCCTGCGAGTTGGAGGCTTCTTCTTCTTCTTTTTTCTTCTTTTTTTTTTTTGGATAACAGA
GTCTCGCTCTTTTTTGCCAGGCTGGAGTGGTTTGGCGTGATCTTGGCTCACTGCAACCTGTGCTTCTGAGTTCAAGCA
ATTCTCTTGCCTCAGCCTCCCAAGTAGCTGGAATTATAGGCGCCACCACCATGCTGACTAATTTTTGTAATTTTAGTAG
25 AGACGAGGTTTCTCATGTTGGCCAGGCTGGTCTCGAACTCCTGACCTCAGGTGATCCTCCACCTCGGCCTCCCAAGT
GCTGGGATGACAGGTGTGAACCGCGCGCCCGCCGAGACTCGCTTCTGAGCTTCCGTGAGATCTGACCGATAGCTG
CCTGCAGCCTTGGTGTGACAACTCCGTTTCTTCTCCAGGTCTCGTAGGGGTCTTTCATTTTCATGACTCTCTTCA
CAGAAGAGTTTACGTGTGCTGATTTCCCGCTGTTTCTGCGTAATGGTGTCTGCTGTTTATCGATGGCCTCCTTCCA
TTTCTTTAGGCTTTGTTTATTGTTGTTTTCCGGCTCCTTGAAGGAAAAGTTTCGATTATGGATGTTTGAACCTTCTTT
30 TCTAAACAAGCATCTGAAGTTGCCGTTTTCCCTCTAAAGCAGGGATCCCGAGGCCCTGGCTGTGGAGTGGCACCGGTCT
GGGGCTGTTAGGAACCGCGCGCAGCGGGAGGCTAGGTGGGGTGTGGGGAGCCAGCGTCCCGCCTGAGCCCCGCCCC
TCTCAGATCAGCAGTGGCATGCGGTGCTCAGAGGCGCACACCCCTACTGAGAACTGTGCGTGAGAGGGGTCTAGATTCT
GTGCTCCTTATGGGAATCTAATGCCTGATGATCTGAGGTGGAACCGTTGCTCCCAAACCATCCCTTCCCCACTGCTG
TCCTGTGGAAAAATCGTCTTCCACGAAACAGTCCCTGGTACCACAATGGTTGGGGACCTGTGCTAAAGACCTGCTTCA
35 GCAGCCTCTCGTCAGTGTGATATATTGGCTTTTCTGTGTTGAGTCCAGAATAATTACGGATTTCTGTGATGCTTCCGC
CGACCTCAGACCCATGGGCTATTTGTGGCGTGTTCCTGCTCCTGGGTTGGGAAGGGTGCAGGCCCATGTACCTTCTCT
GTTACTGCCTTCCAGGTTGGTTCTCAGGTTGAATCGTACTCGATGTGGTTTTAGCCCACGGCCCTGCCGCCAGCTCCTG
GGGGCTGGGGAACATGCTGAAGCACAGAGTCACCGTGCCTGCTTTTGTGCTCACAAGCTCGAGGCTCCTGTGCTCG
TGTTAGTGTGTGTACGTGCTGCTCACATCTGTCTTGGGGACGAGGGGCTTAGCAGGTCCCGTAGTAAATGACAAGC
40 GTCCTGGGGAGTCTGCAGAATAGGAGGTGGGGGTGCCGCTCTCTCCCGCTTCTCAGACTTCTCCTGCTGTGCT

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGCCCGGAGCTCGAGTGCCACTTGTGCCACGT
GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTGGTACAGGGGTCTGATGTGTG
GTGACTGTGGATGGCGGTCGTGGGCTGATGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGG
ATGGCGGTCGTGGGCTCTGATGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGGTGAAGTGTGGATGGCGGTCGTG
5 GGGTCTGATGTGGTGAAGTGTGGATGGCAGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATG
TGGTGAAGTGTGGATGGCAGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAG
GTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGG
CGGTCGTGGGCTCTGATGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTCA
CAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTCA
10 GGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTG
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTG
GATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGGT
GACTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCAG
15 TCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGG
TCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGT
GGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGTATCGGTCAAGGGGTCTGATGTGTGGT
GACTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGAT
GGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCAGTGG
20 GTCAAGGGGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGG
GGTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGATGTGTGGTGAAGTGTGGATGGCGGTCGTGGGCTCTGAT
GTGGTGAAGTGTGGATGGTATCGGTCAAGGGGTCTGATGTGTGGTGAAGTGTGGATGGAGTCCAGGTGTGTCTGTAGCT
ACTTTGCGCTCCTCGGCCCCCGGCCCCGTTTCCCAAAACAGAGCTTCCAGGCGCTCTCTGGGCTTCATCCCGCATCG
GGCTTGGCCGACAGTCCACAGTCTGATCGGAAGAAACAGTGCACAGCTCTGGCCGGGACAGGCCATTTGTGGCTC
25 ATGCCCTCTCCTCTGCCGGCAG

Intron 7 (SEQ ID NO 11)

GTCTGGGCACTGCCCTGCAGGTTGGGCACGGAAGTCCAGCAGTGGGTCCTCCCCTGGGCAATCACTGGGCTCATGACCG
GACAGACTGTTGGCCCTGGGGGGCAGTGGGGGAATGAGCTGTGATGGGGGCATGATGAGCTGTGTGCCTTGGCGAAATC
30 TGAGCTGGGCCATGCCAGGCTGCGACAGTGTGCAATTCAGGCACCTGCTCACGTTGACTGCGCGGCTCTCTCCAGTT
CCGCACTGCCTTTGTTTATGATTTGCTAAATGCTCTCTCTGCCAGTTTGTATCTTGAGGCCAAAGGAAAGGTGTCCCCCT
CCTTTAGGAGGGCAGGCCATGTTTGAAGCGTGTCTGCGCCAGTGGCCCTCAGTGCTGGGTCTGAGGCCAAAGGAAACG
TGTCCCCCTTCTTAGGAGGACGGGCGGTGTTTGAAGCCAGCCCCGCTGAGCGGGCTCTCAGTGCTGGGTCTGTCCACGT
GGCCCTGTGGCCCTTTGAGATGTGGTCTGTCCACGTGGCCCTGTGGCTCTTTGAGATGCTGTAGCACTGTGCTGGC
35 TCTAGGGACAGTCTGTCCACCGCATGAGGCTCAGAGACCTCTGGCGAATTTCTTGGCTCCAGGGTGGGGGTGGAG
GTGGCTGGGCTGTGGGACCCAGACCTGTGCCCGGACGTGGGCAGCAACTCTGGATCACATATGCCATCCGGGCCA
CGGTGGGCTGTGTGGGTGTGAGCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTACACACTCTGCCTAA
GCCCATGTGTGTCTGACAGAGACTCGGCCCGGCCAGCCACAGATGGCCCTGCAATTCAGCCAGCCCGCACTTCATCACA
AACACTGACCCCAAAGGGACGGAGGGTCTTGGCCACGTGGTCTGCTGTCTCAGCACCCACCGGCTCACTCCCATGTG
40 TCTCCGCTGTGCTTTCGAC

5

10

15

20

25

30

35

40

GTGAGGCCTCCTCTTCCCCAGGGGGGCTTGGGTGGGGGTGATTGTCTTTTGATGCATTCAGTGTTAATATTCTGGTGCTCTGGAGACCATGACTGCTCTGTCTTGAGGAACAGACAAGGTGTCAGCCCCCTCTTGGTATGAAGCCGCACGGGAGGGGTTGCACAGCCTGAGGACTGCGGGCTCCACGCAAGGCTCTGTCCAGCGGCCATGTCCAGAGGCCTCAGGGCTCAGCAGGCGGGGGCCGCTGCCCTGCATGATGAGCATGTGAATTCAACACCGAGGAAGCACACCAAGCTTCTGTCACGTCACCCAGGTTCT

CGTTAGGGTCCTTGGGGAGATGGGGCTGGTGAGCCTGAGGCCCCACATCTCCAGCAGGCCCTCGACAGGTGGCCTGGA
CTGGGCGCCTCTTCAGCCCATTTGCCATCCCACTTGTCATGGGGTCTACACCAAGGACGCACACCTAAATATCGTGCC
AACCTAATGTGGTTCAACTCAGCTGGCTTTATTGACAGCAGTTACTTTTTTTTTTTAATACTTTAAGTTCTAGGGTAC
ATGTGCACGACGTGCAGGTTAGTTACATATGTATACATGTGCCATGTTGGTGTGCTGCACCCATTAACTCATCATTTACA
5 TTAGGTATATCTCCTAATGCTATCCCTCCCCACTCCCCCATCCCATGACAGGCCCTGGTGTGTGATGTTCCCCACCCCTG
TGTCCAAGTGTTCTCATTGTTCAAGTCCCACTGTGAGTGAGAACATGTGGTGTGTTGGTTTTCTTTCTTGCATAGTTT
GCTCAGAGTGATGGTTTCCAGCTTCGTCCATGTCCCTACAAAGGACATGAACTCATCCTTTTTTATGACTGCATAGTATT
CCGTGGTGTATATGTGCCACATTTTCTTAATCCAGTCTATCATCGATGGACATTTGGGTGGTTGCAAGTCTTTGCTACT
GTGAATAGTGCCGCAATAAACATACGTGTGCATGTCTTTATAGCAGCATGATTATAATCCTTTGGGTATATACCCAG
10 TAATGGGATGGCTGGGTCAAATGGTATTCTAGTTCTAGATCCTTGAGGAATCACCACACTGTCTTCCACAATGGTTGAA
CTAGTTTACACTCCCACCAACAGTGTAAGAGTTCGTGGTGTGAGAGGATGTGGACAGCAGTTATTTTTTATGAAAA
TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACACAGCCATTTCTCTCGAAGAC
TCCGGGTTTTCTGTGCATCTTTGAACTCTAGCTCCAATTATAGCATGTACAGTGGATCAAGGTTCTTCTTCATTAA
GGTTCAAGTTCTAGATTGAAATAAGTTTATGTAACAGAAAACAAAATTTCTTGATACACAACTTGCTCTGGGATTGGA
15 GGAAAGTGTCCTCGAGCTGGCGGCACACTGGTCAGCCCTCTGGGACAGGATACCTCTGGCCCATGGTCATGGGCGCTGG
GCTTGGGCTGAGGGTCACACAGTGACCATGCCCAGCTTCCTGTGGATAGGATCTGGGTCTCGGATCATGCTGAGGACC
ACAGCTGCCATGCTGGTAAAGGGCACCAGTGGCTCAGAGGGGGCAGGTTCCAGCCCCAGCTTTCTTACCGTCTTCAG
TTATTTTTCCCTAAGAGTCTGAGAAGTGGGGCCGCGCTGATGGCCTTCGTTCTGCTTCAGCTGGCACAGAATTGCACAA
GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACTGTAGAGAGCTCGTCTGTGGAAA
20 GAAATTTAAGTTTTTCATTTAACCGCTTTGGAGAATGTTACTTTATTTATGGCTGTGTAATTTGTGACATTAGTCCC
TCGTAGACAGATACTACGTAAAAAGTGAAGTTAACCTTGCTGTGATTTTCCCTTATTTAG

Intron 10 (SEQ ID NO 14)

GTGAGGCCCCGTGCCGTGTGTCTGTGGGACCTCCACAGCCTGTGGGCTTTGCAGTTGAGCCCCCGTGTCTGCTGCCCTGG
25 CACCGCAGCGTTGTCTCTGCCAAGTCTCTCTCTGCGCGTGCTGGATCCGCAAGAGCAGAGGCGCTTGGCCGTGCACC
CAGGCCTGGGGCGCAGGGGCACCTTCGGGAGGAGTGGGTACCGTGCAGGCCCTGGTCTGCAGAGACGCACCCAGGTT
ACACACGTGGTGAGTGCAGGCGGTGACCTGGCTCCTGCTGCTCTTTGAAAGTCAAGAGTGGCGGCTCCTGGGGCCCCAG
TGAGACCCCCAGGAGCTGTGTCAGGGCCTGCAGGGCCGAGGCGGCAGCCTCCTCCCAGGGTGCACCTGAGCCTGCGGA
GAGCAGGAGCTGCTGAGTGAAGTGGCCACAGCGTTCGCTGCGGTACGTTCTGCTGCGGGGTTGTTTGGGATCGGTGGG
30 AGAATTTGGATTGCTGAGTGCTGCTGCTTGAACCAAGGAGATGGCTAGGAGTGGGTTTTCAGAGTTGATTTTTGTGAAT
CAAACTAAAATCAGGCACAGGGGACCTGGCCTCAGCACAGGGGATTGTCCAATGTGGTCCCCCTCAAGGGCGCCCCACAG
AGCCGGTGGGCTTGTTTTAAAGTGGATTGACGAGGGACGAGAAACCTTGAAAGCTGTAAAGGGAACCTCAGAAAATG
TGGCCGCCAGGGGTGGTTTCAGGTGCTTTGCTGGGCTGTGTTTGTGAAAACCCATTGAGCCCGCCCTCCAAGTCCACCC
TCCAGGTCCACCCTCCAGGGCCGCGCTGGGCTGGGGGTATGCTGGCGTTCTTGTGCGCGAGCCCGAGCACAGCAGGC
35 TGTGCACATTTAAATCCACTAAGATTCACTCGGGGGAGCCAGGTCCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT
GCTGAGGGGACAGAGCAGACGGGAACGCTGCTTCTGTGTGGCAAGTTCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA
GTGTATGTTGGGGTCCACCGGGGGCAGAACTCTGTCTCTGATGAGTCGGCAGCCATGTAAACAGGAAGGGGTGGCCACAG
GGAGCTGGGAATGCACAGGGGAGCTGCGCAGCTGGCCGAGGTCCAGGGCCAGGCCACAGGAAGGGCAGGGGGACGCCC
GGGGCCACAGCAGAGGCGCGAGGAAGGGAAGGGATGCCAGGCCAGAGCAGAGGCTACCGGGCACAGGGGGGCTCCCTG
40 AGCTGGGTGAGCGAGGCTCATGACTCGGCGAGGGAACCTCCTTGACGTGAAGCTGACGACTGGTGTGCCCCAGCTCACAG

CCCAGCCAGGTCCCGCGCCTGAGCAGGAACTCAGAACCCTCCCTTTGTCTAAAGCACAGCAGATGCCTTCAGGGCATCT
AGGAGAAAACAGGCCAAAGTCGTTGAGAAACGTCTTAAAGAAGGTGGGATGGTGGCAATTCTTGTCAGATTTTAGTCT
GCCCCGGACCACAGATGAGTCTATAACGGGATTGTGGTGTGCCATGGGGACACATGAGATGGACCATCACAGAGGCCAC
TGGGGCTGCACCTCCCATCTGAGTCTGGCTGTCCCGGGTCCAGGCCAGGTTCTTGCTAGCTCACCTACCTGTCTGCCCC
5 GGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGGTCCAGGCTCCTCAGAGCTCCTGCCAGGCCAGCACCTT
GCTCCAAATCACCATTCTCTGGGGTTTTCCAAAGCATTTAAACAAGGGTGTGAGGTTACCTCCTGGGTGACGGCCCCGCA
TCCTGGGGCTGACATTGCCCTCTGCCTTAG

Intron 11 (SEQ ID NO 15)

10 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGGTGGGGCAGGTGCTGCTGCAGGGCCGTTGCGTCCACCTCTGCT
TCCGTGTGGGGCAGGCGACTGCCAATCCCAAGGGTCAGAGGCCACAGGGTGCCCTCGTCCCATCTGGGGCTGAGCAGA
AATGCATCTTTCTGTGGGAGTGAGGGTGTCTACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAAAGGAAATGGTGAC
CAGACCTGGGTGCACTGAGGTGTCTTCAGAAAGCAGTCTGGATCCGAACCCAAGACGCCCGGGCCCTGCTGGGCGTGAGT
CTCTCAACCCGAACACAGGGGCCCTGTGGGCATGAGTCCCTCTGAACCCGAGACCCTGGGGCCCTGCTGGGCGTGAGT
15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTTTGGGCGTGAGTCTCTCCGTGTGAGCCCCACACTCCAAGGCTCATCCAC
AGTCTACAGGATGCCATGAGTTCATGATCACGTGTGACCCATCAGGGGACAGGGCCATGGTGTGGGGGGGTCTCTACAA
AATTCTGGGGTCTGTTTCCCAGAGCCCGAGAGCTCAAGGCCCGTCTCAGGCTCAGACACAAATGAATTGAAGATGGA
CACAGATGCAGAAATCTGTGCTGTTCTTTTATGAATAAAAAGTATCAACATTCAGGCAGGGCAAGGTGGCTCACACCT
ATAATCCAGCACTTTGGGAGGCCGAGGTGGTGGATCACTTGAGGCCAGGAGTTTGAGGCCAACCTAACCAACATAGTG
20 AAATTCATTTCTACTTAAAAAATACAAAAATTAGCCTGGCCTGGTGGCACACGCTGTAGTCCCCGCTATGCGGGAGGC
TGAGGCAGGAGAATCATTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGAGATCACACCACTGCCTCCAGCCTGGGCA
ACAGAGTGAGACTTCATCTTAAAAAAGTATCAGCATTCCAAAACCATAGTGAGCAGGTGTTTTTTTATTC
TGTCTTCGATAATATTACTGGTGTGTGCTAGAGGCCGGAACCTGGGGTGCCCTTCTCTGAAAGGCACACCTTCATGG
GAAGAGAAATAAGTGGTGAATGGTTGTTAAACCAGAGGTTTAAACTGGGGTCTGTGTTCTGAGTTAACAGTCCAGATC
25 TGGACTTTGCCTCTTCCAGAATGCTCCCTGGGGTTTGCTTCATGGGGAGCAGCAGGTGTGGACACCCTCGTGATGGGG
GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCCTTGTCATGGTGCCAGCATGTCCCTG
TTGAGCTCCCTCCCCACAAGGATGCCGGTCTCCTGTGCTCCCCACAGTCCCTGCTTCCCTCTCACAGCCTTACCTGGTC
CTGGCCTCCACTGGCTTTGTCTGCATGATTTCCACATTTCTGGGGTCCCAGCACCTCTTCGCTCTCCAGGCACCTCT
GCAGTGTGCCCATAACAGTCAGCTGTGAACGTCCACTGCTTATTTTGTCTCCCATGAAATGTATTTTTTAGGACAGGC
30 ACCCTGTGTTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAAATGG
GTTCTCTCTAAACACATTGCAAGCCACAGAGGCTAGTGCAGGATGGGTGGGCATCAGGTCTCAGATGTGGGTCCAATG
CCAGAATATTCTGTGCTCCCAAGGCCACTTGGTCAGAGTGTGTGCTTGAGAGGTGGCTCTAAAGCTCAGCAGTGGAG
GCAGTGGTTGCGCATACTCAGGGTGAACCTCACATCCTCTGTGTCTGAAGTATACAGCAGAGGCTTGAAGGGCATCTGGGA
GAAGAAAACAGGCAAAATGATTAAGAAAAGTGAAGGAAAGTGGTAAGATGGGAATTTCTTGTCAGATTTTAGTCTC
35 TCCCAACCAAGCTCAGATGGTAGAATGTGGTCAAGTGTGACAGAAACATAGAACAAAACGGAAGCCCTATCTCT
CAGAAACGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAGAGAGTGTGTGTGAATTTTTTTTCTGAGAAAAC
GACTGGAAGCAAATAAGTTGTGCTTTACAGCATATACCAGAGCAGATTCTAGGTAGAAGAGGAGACATGCAAAACAC
ACCAGCAACAGAAATAAAACAAAGACTCAAAGGGAAGGGAGGTGAACGTTCCCTGGTTTGGTGTGGGGAAGGACACAC
AGGGAGGCGGATGAAACAGTGAGGCAACGGGCATTGCTTTCACTGCAGAGAAAACCTCAGCTTGCTGAGCCACAGTGAAA
40 ATGGCCATTCCCTGGAGCGTTTGTGACGTGATTTATTAAAGGCGCCCTGTGAGGTCTGCACATTATCCTCTCACTTT

GTTCCTCTAACCACTGAGAGGTAGAGGAGAAAGGCTCCAGGGGAGCAGCCGCCCTTGGTCACCCAGCTGGCAAAGGGC
ATGCATGATTGCAGCCTGGCCTCCTGCTCCGGGGCCCTTGCTCTGCCCCGAGGACCCACACAAGTCAGACCCATAGGCTC
AGGGTGAGCCGGAGCCCAAGGTCTGTGTTGGGGATGGCTGTGAAAGAAGAAATGGACGTCTGATGCACACTTGGGAAGGTC
CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCAGTGAAACTGATGGC
5 GAGACCTGTCCCCATCCCTCATGCTGGCTCCTTTCTGGGCTTGCCAAGAGCCAGCATCAGGTTGAGGCAAGCTGGAAAG
ACTTTTCTGGAAAGCAGCTTGTGTCATGGAAGTCTCACAATGTCTGTCTTCCAGTAATTCACCTTCTGAAGTGA
CCAGACATTATCACGGGTCTTATTTACCATTTCAGTGTTCAGGCAGGGGGACTTGCCACAGCAAGTCAGAACCTGCC
CAAATACAGGGCTAAGGAGATATTATGCATCACAAAATTGCTCTGCCATTAAACATTTTCAAAGAATTTTGAAGAAT
GTTTAATGGCACAAAACGTTTATTTCAATGTAGCAGTGTCAAAGCTGGATGTAAAGAACACACCCAGGAGCCTGCCG
10 TGAATGTATGTGTGTTTCTTTGGACATGGACATACATGGGCAGTGAGTGGTGGTGAGGCCCTGGAGGACATCGGTGG
GATGCCTCCATCCTGCCCTCTGGAGACACCATGTGTGCCAGTGCACTCACTGGAGCCCTGTTTAGCTGGTGCCACCTG
GCTCTTCCATCCCTGAGATTCAAACACAGTGAGATTCCCCACGCCCACTCAGTGTCTCCACAAAAAACCTGAGTCAC
ACCTGTGTTCACTCGAGGGACGCCGGGAGCCAGGGCTCCACAGTTTATTATGTGTTTTGGCTGAGTTATGTGCAGATC
TCATCAGGGCAGATGATGAGTGCACAAACACGGCCGTGCGAGGTTTGGATACACTCAACATCACTAGCCAGGTCCTGGTG
15 GAGTTTGGTCATGCAGAGTCTGGATGGCATGTAGCATTGAGAGTCCATGGAGTGAGCACCAGCCCCCTCGGGCTGCAGC
GCATGCCCCAGGCAGGACAAGGAAGCGGGAGGAAGGCAGGAGGCTCTTTGGAGCAAGCTTTGAGGAGGGGGCTGGGTGT
GGGGCAGGCACCTGTGTCTGACATTCCCCCTGTGTCTCAG

Intron 12 (SEQ ID NO 16)

20 GTGAGCAGGCTGATGGTCAGCAGAGTTCAGAGTTTCAAGAGGTGTGTGCGCAAGTATGTGTGTGTGTGTGCGCGCT
GCCTGCAAGGCTGATGGTGACTGGCTGCACGTAAGAGTGACATGTACGCATATACAGTGAGCAGATACATGTGTGCAT
GTGTGTACATGAAGGCATGGCAGTGTGTGCACAGGTGTGCAAGGGCACAAGTGTGTGCATGCGAATGCACACCTGACA
TGATGTGTGTTCTGTGCAGTCTGTGGGCATTACGTGAGGTGCATGCGTGTGGGTGTGCAGTGTGAGTAGCATGTGT
GCACATAACATGTATTGAGGGGTCTCGTGTTCACCCCGCTAGGTCTCAGCACCAGTGCCACTCCTTACAGGATGAGAC
25 GGGGTCCAGGCCTTGGTGGGCTGAGGCTCTGAAGCTGCAGCCCTGAGGGCATTGTCCACTCTGGGCATCCGCGTCCACT
CCCTCTCCTGTGGGCTTCTGTGTCCACTCCCCCTCTCCTGTGGGCATTACATCCACTCCACTCCCTCTCTCCTGTGGGC
ATCCGCGTCCACTCCCCCTCTCTGTGGGCATCTGCGTCCACCTCCCCCTCTCTGTGGGCATTGCGTCCACTCCCTCTCCT
GGTTCCTTCTGTCTTGGCCGAGCCTCGGGGCGAGGCAGATGACACAGAGTCTTGACTCGCCAGGGTGGTTCGCAAGCTG
CCGGGTGAGGGCCAGGCCGATTTCCTGGAAGAGGGATAGTTTCTGTCAAATGTTCTCTTCTTGTTCATCTGA
30 ATGGATGATAAAGCAAAAAGTAAAACTTAAATCCAGAGAGGTTTCTACCGTTTCTCACTCTTCTTGGCGACTCTAG

Intron 13 (SEQ ID NO 17)

GTGAGCCGCCACCAAGGGGTGCAGGCCAGCCTCCAGGGACCTCCGCGCTCTGCTCACCTCTGACCCGGGGCTTCACTT
TGGAACCTCTGGGTTTTAGGGGCAAGGAATGTCTTACGTTTTTCAAGTGGTGTCTGTGCTGTGCACAGTCTGTTCGCGTG
35 GCTCTGTGCAAAGCACCTGTCTCCATCTCTGGGTAGTGGTAGGAGCCGGTGTGGCCCCAGGTGTCCCCACTGTGCTGT
GCACTGGCCGTGGGACGTATGGAGGCCATCCAGGGCAGCAGGGGCATGGGGTAAAGAGATGTTTATGGGGAGTCTTAG
CAGAGGAGGCTGGGAAGGTGTCTGAACAGTAGATGGGAGATCAGATGCCCGAGGATTTGGGGTCTCAGCAAAGAGGGCC
GAGGTGGGTGCAGGTGAGGGTCGTGGCCCCACCCCGGAAGGTGCAGCAGAGCTGTGGCTCCCCACACAGCCCGCCA
GCACCTGTGCTCTGGGCATGGCTGTGCTCCTGGAACGTTCCCTGTCTGGCTGGTCAGGGGGTGCCCTGCCAAGAATCG
40 ACAACTTTATCACAGAGGAAGGGCCAATCTGTGAGGCCACAGGGCCAGCTTCTGCTGGAGTCAGGGCAGGTGGTGGC

ACAAGCCTCGGGGCTGTACCAAAGGGCAGTCGGGCACACAGGCCCGGGCCTCCACCTCAACAGGCCTCCCGAGCCACTG
GGAGCTGAATGCCAGGAGGCCGAAGCCCTCGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTTGTGTTACCCAGGCGCG
AGGCTGCGCGAATTACCGTGCACACTTGATGTGAAATGAGGTCGTCTATCGTGGAACCCAGCAAGGGCTCACGGGA
GAGTTTTCCATTACAAGGTCGTACCATGAAATGGTTTTTAACCCGAGTGCTTGCGCCTTCATGCTCTGGCAGGAGGGG
5 AGAGCCACAGCTGCATGTTACCGCCTTTGCACCAAGCTCCAGAGGCTTGGGACCAGGCTGTCTCAGTTCCAGGGTGCGTCC
GGCTCAGACCGCCCTCCTCTGCTCTCTCTCTGCTCAAAATCTTCCCTCGTTTGCATCTCCCTGACGCGTGCTGGG
CCCTCGTGCAAGCTGCTTGACTCCTTTCCGAAACCCCTTGGGGTGTGCTGGATACAGGTGCCACTGAGGACTGGAGGTGT
CTGACACTGTGGTTGACCCAGGGTCCAGCTGGCGTGTCTGGGGCCTCCTGGGCCATGATGAGGTGAGGAGAGTTTCC
CAGGTGAAAACCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGTCTCTCCCATATTCAGCTCAGTCTTGTCTCTC
10 ATTTCCCAACAGGGTCTCTAGCTCCGAGGAGCTCCCGTAGAGGGCCTGGGCTCAGGGCAGGGCGGCTGAGTTTCCCAAC
CCATGTGGGGACCCCTTGGGTAGTCGCTTGATTGGGTAGCCCTGAGGAGGCCGAGATGCGATGGGCCACGGGCCGTTTCCA
AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCTTCCCTCGAGGCAGGAGTGGGAGAACGAGAGCTGGGCCCGG
ATTTACGGCAGCCAGGCTGCAGTGGGCGAGGCTGTGGTGGTCCACGTGGCGCTGGGGCGGGGTCTGATTCAAATCCGC
TGGGGCTCGGCCTTCTGGCCCGTGTGGCCGCGCTCCACACGGGCTTGGGGTGGACGCCCCGACCTCTAGCAGGTGGC
15 TATTTCTCCCTTTGGAAGAGAGCCCTCACCATGCTAGGTGTTTCCCTCCTGGGTGAGGAGCGTGGCCGTGTGGCAACC
CCGGGACCTTAGGCTTATTTATTTGTTTAAAAACATTCTGGGCTGGCTTCGTTGTTGCTAAATGGGGAAAAGACATCC
CACCTCAGCAGAGTTACTGAGAGGCTGAAACCGGGTGTGGCTTACTGGTGTGATCTCAGGTCAATCCAGAAGTGGCT
CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGTGGGT
GAGGCCAGGTACATGGGGGCTCAGGCAGTGGGTGAGATGAGGTACACGGGGGCTCAGGCAGAGGCTCAGACCAGGTAC
20 ACGGGGCTCTGATCACACGCACATATGAGCACATGTGCACATGTGCTGTTTCATGGTAGCCAGGTCTGTGCACACCTGC
CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGAGGGCTGACAGGGCTGGCGCGGTGGCTCACACCTGTAGTCCCAGCA
CTTTGGGAGGCCGAGGCGAGAGGATCCCTTGAGCCAGGAGTTTAAGACCAGCCTGAGCAACATAGTAGAACCCTATCTC
TATGAAAAATAAAACAAAAATTAGCTGAACATGGTGGTGTGCGCCTGTAGTTCCAATACTTGGGAGGCTGAAGTGGGAG
GATCACTTGAGCCAGGAGGTGGAAGCTGCAGTGAGCTGAGATTGCACCACTGTACTGACGCTGGGTGACAGAGTGAGA
25 GCCCATCTCAACAACAACAAGAAGACTGACAAATGCAGTTTCTTGGAAGAAACATTTAGTAGGAACTTAACCTACACA
CAGAAGCCAAGTCGGTGTCTCGGTGTCTGAGATGAGATGAGGTCTCACACCATCACCCAGACCCAGGGTTTATG
CACCACAGGGGCGGTGGCTCAGAAGGGATGCGCAGGACGTTGATATACGATGACATCAAGGTTGTCTGACGAAGGGCAG
GATTCAAGATAAGTACCTGCTGGTACACAAGGAACAATGGATAAACTGGAAACCTTAGAGGCCTTCCCGAACAGGGGCT
AATCAGAAGCCAGCATGGGGGCTGGCATCCAGGATGGAGCTGCTTCAAGCTCCACATGCGTGTTCATACAGATGGTGCA
30 CAGAAACGCAGGTGACCTGTGCACACAGACACGAGCTACTCGCACACACAAGCACACACAGACATGCATGCATGC
ATCCGTGTGTGTGCACCTGTGCCATGAGGAAACCATGATGTGCATTATGCACGCACACAGGCACCGGTGGGCCCAT
GCCACACCCACGAGCACCCTCTGATTAGGAGGCTTCTCTGACGCTGTCCGCCATCTCTCAG

Intron 14 (SEQ ID NO 18)

35 GTATGTGACAGGTGCCTGGCCTCAGTGCCAGCAGTGCCCTGCTGCTGGTGTAGTGTGTCAGGAGACTGAGTGAATCTGGG
CTTAGGAAGTTCTTACCCCTTTTCGCATCAGGAAGTGGTTTAACCAACCACTGTGAGGCTCGTCTGCCCGCCCTCTCGT
GGGGTGAGCAGAGCACCTGATGGAAGGACAGGAGCTGTCTGGGAGCTGCCATCCTTCCACCTTGCTCTGCTGGGGAA
GCGCTGGGGGCGCTGCTCTCTGTTTGGCCATGTTGGGATTGGGGGGCCTGGCCTCTCTGTTTGGCCCTGTGGTGG
GATTGGGCTGTCTCCGTCATGGCACTTAGGGCCCTGTGTCAAACCCAGGCCAAGGGCTTAGGAGGAGGCCAGGCCAG
40 GCTACCCACCCCTCTCAGGAGCAGAGGCCGCTATCACACGACAGAGCCCCGCGCGTCTCTGCTTCCAGTACCCG

TCCTCTGCCCCCTGGACACTTTGTCCAGCATCAGGGAGGTTTCTGATCCGTCTGAAATTCAAGCCATGTCGAACCTGCGGT
CCTGAGCTTAACAGCTTCTACTTTCTGTTCTTCTGTGTTGTGGAAATTTACCTGGAGAAGCCGAAGAAAACATTTCTG
TCGTGACTCCTGCGGTGCTTGGGTGCGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCCG
GTGTCTCCTGGGAGGGGAGCTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG

5

Intron 15 (WEQ ID NO 19)

GCAAGTGTGGGTGGAGGCCAGTGCGGGCCCCACCTGCCAGGGGTATCCTTGAACGCCCTGTGTGGGGCGAGCAGCCTC
AGATGCTGCTGAAGTGCAGACGCCCCGGGCTGACCCTGGGGGCTGGAGCCACGCTGGCAGCCCTATGTGATTAAACG
CTGGTGTCCCCAGGCCACGGAGCCTGGCAGGGTCCCCAACTTCTTGAACCCCTGCTTCCCATCTCAGGGGCGATGGCTCC
CCACGCTTGGGAGCCTTCTGACCCCTGACCTGTGTCTCTCACAGCCTCTTCCCTGGCTGCTGCCCTGAGCTCCTGGGGT
CCTGAGCAAGTTCTCTCCCGCCCCCGCTCCAGCGTCACTGGGCTGCTGTCTGCTCGCCCCGGTGGAGGGGTGTCTG
TCCCTTCACTGAGGTTCCACCAGCCAGGGCCACGAGGTGCAGGCCCTGCTGCCCCGCCACCCACGCTCCTAGGAGGG
TTGGAGGATGCCACCTCTGGCCTCTTCTGGAACGGAGTCTGATTTTGGCCCCGAG

10

3'-untranskribierter Bereich (SEQ ID NO 20)

ATCTCATGTTGAATCCTAATGTGCACTGCATAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGGTGGT
GGTCAGTGCGGGCCCATGGCCTGGCTGTGCATTTACGGAAGTCTATGAGTGAATGGGGTTGTGGTCAGTGCGGGCCCATG
GCCTGGCTGGGCCTGGGAGGTTTCTGATGCTGTGAGGCAGGAGGGGAAGGAGGGTAGGGGATAGACAGTGGGAGCCCCCA
CCCTGGAAGACATAACAGTAAGTCCAGGCCGAAGGGCAGCAGGGATGCTGGGGGCCAGCTTGGGCGGCGGGATGATG
GAGGGCCTGGCCAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGGGCTGGTCTGGGTGG
CGGGGAAGATGGGAAGCCTGGCTGGGCCCCCTCCTCCCTGCCTCCACCTGCAGCCGTGGATCCGGATGTGCTTCCCT
GGTGACATCCTCTGGCCATCAGCTTTTCTGAGGTGGGGGGCAGGGGCATGACACCATCCTGTATAAAATCCAGGATT
CCTCCTCCTGAACGCCCAACTCAGGTTGAAAGTCACATTCGCTCTGGCCATTCTCTTAAGAGTAGACCAGGATTCTG
ATCTCTGAAGGGTGGGTAGGGTGGGGCAGTGGAGGGTGTGGACACAGGAGGCTTCAGGGTGGGGCTGGTGATGCTCTCTC
ATCCTCTTATCATCTCCAGTCTCATCTCTCATCTCTTATCATCTCCAGTCTCATCTGTCTTCTCTTATCTCCAGT
CTCATCTGTCTCTTACCATCTCCAGTCTCATCTCTTATCCTCTTATCTCCTAGTCTCATCCAGACTTACCTCCCA
GGGCGGGTGCCAGGCTCGCAGTGGAGCTGGACATACGTCCTTCTCAGGCAGAAAGAACTGGAAGGATTGCAGAGAACAG
GAGGGGCGGCTCAGAGGGACGCAGTCTTGGGGTGAAGAAACAGCCCTCCTCAGAAGTTGGCTTGGGCCACACGAACCG
AGGGCCCTGCGTGAGTGGCTCCAGAGCCTTCCAGCAGGTCCCTGGTGGGGCCTTATGGTATGGCCGGGTCTACTGAGTG
CACCTTGGACAGGGCTTCTGGTTTGTAGTGACCCGGACGTGCCTGGTGTGGGGTGGGGCTTATGGCCACTGGATATG
GCGTCATTTATGCTGCTGCTTCAGAGAATGTCTGAGTGACCGAGCCTAATGTGTATGGTGGGCCCAAGTCCACAGACTG
TGTGCTAAATGCACTCTGGTGCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGGCTCTTGGCAGCCGGCCTGGGG
GCGCCTTTGCCCTGCAAACTGGAAGGGAGCGGCCCGGGCGCGTGGGCGGACGACCTCAAGTGAGAGGTTGGACAGAAC
AGGGCGGGGACTTCCAGGAGCAGAGGCCGCTGTCTCAGGCACACCTGGGTTTGAATCACAGACCAACaGGTCAGGCCATT
GTTGAGCTATCCATCTTCTACAAAGCTCCAGATTCTGTTTCTCCGGGTGTTTTTGTGTGAATTTTACTCAGGATTACT
TATATTTTTTGCTAAAGTATTAGACCCTTAAAAAAGGTATTGCTTTGATATGGCTTAACTCACTAAGCACCTACTTTAT
TTGTCTGTTTTTATTATTATTATTATTATTATTAGAGATGGTGTCTACTCTGTCAACCAGGTTGTAGTGAGTGGCAC
AGTCATGGCTCGCTGTAGCCGAAACCCAGGCTCAAGTGATCCTCCGGCCTCAGCTTCCAGAGTGTGGGATTACAG
GTGTGAGCCACTGCCCTTGCCTGGCACTTTTAAAAACCACTATGTAAGGTCAGGTCCAGTGGCTTCCACACCTGTCATCC
CAGTAGTTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTTGAGACCAGCATGGGTAACATAGGGAGACCCC

15

20

25

30

35

40

ATCTCTACAAAAATGCAAAAAGTTATCCGGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTCGGGAGGCTGAGTGGG
AGGATCGCTTGAGCCCGGAGGTCATGGCTGCAGTGAGCTGTGATTGTACCATCGCACTCCAGCCTGGGCAACAGAGTGA
GACCCTGTCTCAAAAAAAAAAAAAAAAAAGAGGAGAAGGAGAAGAGAAGAGAAGGAAGGAAGAGAAGAGAAG
5 GAAGAAGGAAGAAAGAGGAGAAGGAGGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAAATGAAAATAACA
AAGTTTAAAGGGAAAGAAAAACCCAGCTCTTTGGACTTCCTTAGGCCTGAACCTCATCTCAAGCAGCTTCCTTCCACA
GACAAGCGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAAGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA
GCCAGGACCCCTGAAAGGGAGTGGTTGTTTTCTGCCTCAGCCCCACGCTCCTGCCGGTCTGCACTGCTGTAACCGTC
GATGTTGGTGCCAGGTGCCACCTGGGAAGGATGCTGTGCAGGGGGCTTGCCAACTTTGGTGGGTTTCAGAAGCCCCAG
10 GCACTTGTGGCAGGCACAATTACAGCCCTCCCAAAGATGCCACGTCCTTCTCCTGGAACCTGTGAATGTGTCACCCG
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTCATCCTGGATTATCTGG
TGGGCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC
CACTGGCCACTGCTGGCTTTGAGATGGAGGAGGGGGTCCCCAGCCAAGGAATGGGGGCAGCCGCTCCATGCTGGAAAAGC
AAGCAATCCTCCCGGTCCTGAGGGCACACGGCCCTGCCACGCCTCGATTTCAGGCCAGTGGGACCTGTTTCAGCTTTC
CGGCCTCCAGAGCTGTAAGATGATGCGTTTGTGTTCAAGCTAAGCTGCAGTGATTTCGTACAGCAGCAAATGGAATAG
15 CAGTACAGGGAATGAATACAGGGACAGTTCTCAGAGTGACTCTCAGCCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorenggeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalisierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, läßt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer internen 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein veränderter C-Terminus resultiert.

Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA⁺-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur
10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv
markierter Primer (5'-GTTAAGTTGTAGCTTACACTGGTTCTC 3'; 2,5-8x10³
cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl
inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL),
2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5 µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl
15 ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser
erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die
Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach
Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden
2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150 µl
20 Phenol/Chloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-
Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min
bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol
gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stopplösung gelöst. Nach 5
min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel
25 aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom
ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in
Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA₋₁TTGT)
30 repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA_n/Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losten Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei
10 anderen TATA-losten Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem
25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer
30 zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer C5S: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCTCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTTCGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCAGGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

Beispiel 8

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

Die Suche nach möglichen Transkriptionsfaktorbindungsstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTTCC-3', Position -4093 bis -4070) wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaares PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGTGGGGATTTCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primerkombination PK-3A und PK-5D (5'-

GGAGATCTGATCTTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmentes (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die
5 verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und
10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 α -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für
15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection
25 bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

30

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

- 5 CHO-K1 und HEK 293 Zellen wurden in wasssergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO₂ kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH₂PO₄ pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit
10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszu-plattieren.

- Für die Transfektion wurden pro Well jeweils 2x 10⁵ –HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium
15 entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serum-freies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium
20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H₃PO₄; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl
25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO₃)₄ Mg(OH)₂·5H₂O; 2,67 mM MgSO₄; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

- 30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysats und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl₂;

50 mM β -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100 μ l 1 M Na_2CO_3 gestoppt und die Absorption bei 420 nm bestimmt.

- 5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotorsequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK
10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im
15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

Literaturverzeichnis

- 5 Allsopp, R. C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.
- 10 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.
- Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.
- 15 Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.
- 20 Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.
- 25 Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol. 14:4894.
- 30 Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.
- 35 Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

- Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.
- 5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.
- 10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.
- Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.
- 15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.
- Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.
- 20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.
- 25 Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.
- Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.
- 30 Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.
- Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.
- 35 Smale, S.T. and Baltimore, D. (1989). The „initiator“ as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5 Shay, J. W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

10

Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

15

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

Patentansprüche

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen
Telomerase-Untereinheit.
5
2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16,
17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser
Sequenzen handelt.
10
3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um
die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen
katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder
um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.
15
4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der
Ansprüche 1 bis 3.
5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß
es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide
oder Proteine kodieren.
20
6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.
- 25 7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem
der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.
8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vek-
toren gemäß einem der Ansprüche 4 bis 6.

30

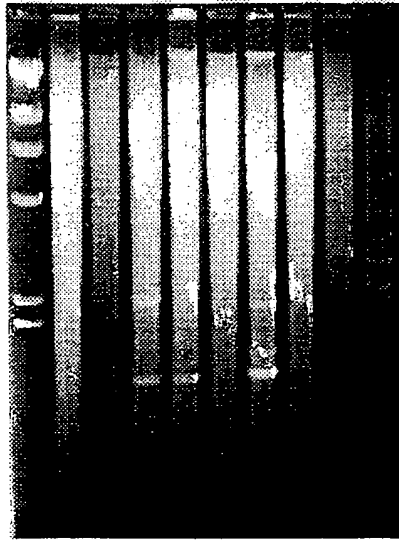
9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:
- 5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,
- B. Messung des Substanzeffektes auf die Reportergenexpression.
- 10 10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screent.
- 15 11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.
- 20 12. Verfahren zur Detektion Telomerase-assoziiierter Zustände bei einem Patienten, das folgende Schritte umfaßt:
- A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält
- 25 mit Körperflüssigkeiten oder zellulären Proben,
- B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

- C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

Fig. 1

A

1 2 3 4 5 6 7 8 9 10



B

1 2 3 4 5 6 7 8 9 10

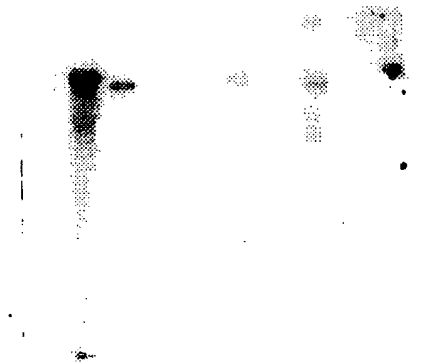


Fig. 2

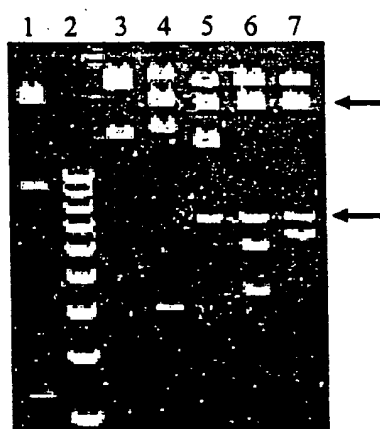


Fig. 3

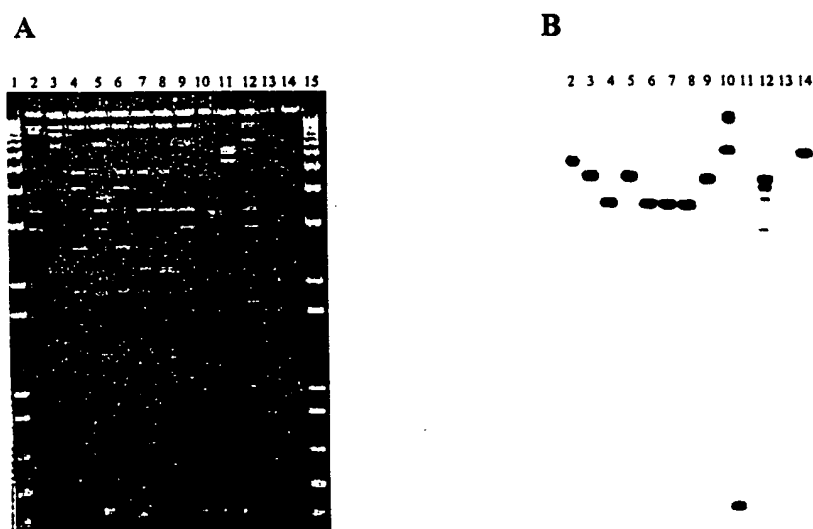


Fig. 4

GAGCTCTGAA	CCGTGGAAAC	GAACATGACC	CTTGCCTGCC	TGCTTCCCTG	GGTGGGTCAA	GGGTAAATGAA	70
GTGGTGTGCA	GGAAATGGCC	ATGTAAATTA	CACGACTCTG	CTGATGGGGA	CCGTTCCTTC	CATCATTATT	140
CATCTTCACC	CCCAAGGACT	GAATGATTCC	AGCAACTTCT	TCGGGTGTGA	CAAGCCATGA	CAAAACTCAG	210
TACAAACACC	ACTCTTTTAC	TAGGCCACCA	GAGCACGGGC	CACACCCTTG	ATATATTAA	AGTCCAGGAG	280
AGATGAGGCT	GCTTTCAGCC	ACCAGGCTGG	GGTGACAACA	GCGGCTGAAC	AGTCTGTTC	CTAGACTAG	350
TAGACCCCTGG	CAGGCACTCC	CCCAAATCT	AGGGCCTGGT	TGCTGCTTCC	CGAGGGCGCC	ATCTGCCCTG	420
GAGACTCAGC	CTGGGGTGCC	ACACTGAGGC	CAGCCCTGTC	TCCACACCCT	CCGCTCCAG	GCCTCAGCTT	490
CTCCAGCAGC	TTCTTAAACC	CTGGGTGGGC	CGTGTTCAG	CGCTACTGTC	TCACCTGTCC	CACTGTGTCT	560
TGCTTCAGCG	ACGTAGCTCG	CACGGTTCCT	CCTCACATGG	GGTGTCTGTC	TCCTTCCCCA	ACACTCACAT	630
GCCTTGAAGG	GAGGAGATTG	TGCGCTCTCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCC	700
GATGCAGGTT	CCTGGCGTCC	GGCTGCACGC	TGACCTCCAT	TTCCAGGCGC	TCCCCTGCTC	CTGTCTATCTG	770
CCGGGGCCTG	CCGGTGTGTT	CTTCTGTTTC	TGTGCTCCTT	TCCACGTCCA	GCTGCGGTGTG	TCTCTGCCCC	840
CTAGGGTCTC	GGGGTTTTTA	TAGGCATAGG	ACGGGGGCGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	910
CATTTGGGTG	TGAAAGTAGG	AGTGCCTGTC	CTCACCTAGG	TCCACGGGCA	CAGGCTGGG	GATGGAGCCC	980
CCGCAGGGA	CCCGCCCTTC	TCTGCCACGC	ACTTTCCTGC	CCCCCTCCCT	CTGGAACACA	GAGTGGCAGT	1050
TTCCACAAGC	ACTAAGCATC	CTCTTCCCAA	AAGACCCAGC	ATTGGCACCC	CTGGACATTT	CCCCACAGC	1120
CCTGGGAATT	CACGTGACTA	CGCACATCAT	GTACACACTC	CCGTCCACGA	CCGACCCCCG	CTGTTTTATT	1190
TTAATAGCTA	CAAAGCAGGG	AAATCCCTGC	TAAAATGTCC	TTTAAACAA	TGGTTAAACA	AACGGGTCCA	1260
TCCGCACGGT	GGACAGTTCC	TCACAGTGAA	GAGGAACATG	CCGTTTATAA	AGCCTGCAGG	CATCTCAAGG	1330
GAATTACGCT	GAGTCAAAAC	TGCCACCTCC	ATGGGATACG	TACGCAACAT	GCTCAAAAAG	AAAGAATTTC	1400
ACCCCATGGC	AGGGGAGTGG	TTAGGGGGGT	TAAGGACGGT	GGGGGCGGCA	GCTGGGGGCT	ACTGCACGCA	1470
CCTTTTACTA	AAGCCAGTTT	CCTGGTTCCTG	ATGGTATTGG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	1540
TGGGATGGGG	GGAAACCCGGA	GGCTGTGCCA	TCTTTGCCAT	GCCCGAGTGT	CCTGGGCGCC	ATAATGCTCT	1610
AGAGATGCCC	ACGTCTCTGAT	TCCCCCAAAC	CTGTGGACAG	AACCCGCCCG	GCCCCAGGGC	CTTGTCCAGT	1680
GTGATCTCCG	TGAGGACCCT	GAGGTCTGGG	ATCCTTCGGG	ACTACCTGCA	GGCCCCAAAA	GTAAATCCAGG	1750
GGTTCTGGGA	AGAGCGCGGC	AGGAGGGTCA	GAGGGGGGCA	CGCTCAGGAC	GATGGAGGCA	CTCAGTCTGA	1820
GGCTGAAAAG	GGAGGGAGGG	CCTCGAGCCC	AGGCCTGCAA	GCGCTCCAG	AAGCTGGAAA	AACCGGGGAA	1890
GGGACCCCTC	ACGGAGCCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	1960
TCCGGCCTCC	GTGCCATAGG	AGGGCACTCG	CGCTGCCCTT	CTAGCATGAA	GTGTGTGGGG	ATTTCAGAAA	2030
GCAACAGGAA	ACCCATGACG	GTGAATCTTA	GGATTATTTC	AAAACAAAGG	TTTACAGAAA	CTGCTCAAGG	2100
CAGGGCTGAA	GTGCTCCCG	GCAAGGGCAG	GGCAGGCACG	AGTGATTTTA	TTTAGCTATT	TTATTTTATT	2170
TACTTACTTT	CTGAGACAGA	GTTATGCTCT	TGTTGCCCAG	GCTGGAGTGC	AGCGGCATGA	TCTTGGCTCA	2240
CTGCACACCT	CGTCTCTGG	GTTCAAGCAA	TTCTCGTGCC	TCAGCCTCCC	AAGTAGCTGG	GATTTCAGGC	2310
GTGCACCACC	ACACCCGGCT	AATTTTGTAT	TTTTAGTAGA	GATGGGCTTT	CACCATGTTG	GTCAAGCTGA	2380
TCTCAAAATC	CTGACCTCAG	GTGATCCGCC	CACCTCAGCC	TCCCAAAGTG	CTGGGATTAC	AGGCATGAGC	2450
CACCTGCACCT	GGCCTATTTA	ACCATTTTAA	AACCTCCCTG	GGCTCAAGTC	ACACCCACTG	GTAAAGAGTT	2520
CATGGAGTTT	AATTTCCCTT	TTACTCAGGA	GTTACCCCTC	TTTGATATTT	TCTGTAAATC	TTCTGAGACT	2590
GGGGATACAC	CGTCTCTTGA	CATATTACAA	GTTTCTGTGA	CCACCTGTTA	TCCCATGGGA	CCCACTGCAG	2660
GGGCAGCTGG	GAGGCTGCAG	GCTTCAGGTC	CCAGTGGGGT	TGCCATCTGC	CAGTAGAAAC	CTGATGTAGA	2730
ATCAGGGGCG	AAGTGTGGAC	ACTGTCCCTA	ATCTCAATGT	CTCAGTGTGT	GCTGAAACAT	GAATCAATTA	2800
AAGTCCATCC	CTCCTACTCT	ACTGGGATTG	AGCCCCCTCC	CTATCCCCC	CCAGGGGCG	AGGAGTTCTT	2870
CTCACTCCTG	TGCAGGAAGG	AATGATACTT	TGTTATTTTT	CACCTGCTGT	ACTGAATCCA	CTGTTTCAAT	2940
TGTTGTGTTG	TTTGTTTTGT	TTTGAGAGGC	GGTTTCACCT	TGTTTGTCTA	GGCTGGAGGG	AGTGCAATGG	3010
CGCGATCTTG	GCTTACTGCA	GCCTCTGCCT	CCCAGGTTCA	AGTGATTCTC	CTGCTTCCCG	CTCCCAATTG	3080
GCTGGGATTA	CAGGCACCCG	CCACCATGCC	CAGCTAAATT	TTTGTATTTT	TAGTAGAGAC	GGGGGTGGGT	3150
GGGGTTCCAC	ATGTTGGCCA	GGCTGGTCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCC	TCTGCCCTCT	3220
AAAGTGCTGG	GATTACAGGT	GTGAGCCACC	ATGCCCAGCT	CAGAAATTAC	TCTGTTTAGA	AACATCTGGG	3290
TCTGAGGTAG	GAAAGCTCACC	CCACTCAAGT	GTTGTGGTGT	TTTAAAGCAA	TGATAGAAAT	TTTTTATTGT	3360
TGTTAGAAC	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTTCAAAG	ACACACTAAC	3430
TGCACCCATA	ATACTGGGGT	GTCTTCTGGG	TATCAGCAAT	CTTCATTGAA	TGCCGGGGAG	GGTTTCCCTG	3500
CCATGCACAT	GGTGTAAATT	ACTCCAGCAT	AATCTTCTGC	TTCCATTCTT	TCTCTTCCCT	CTTTTAAAT	3570
TGTGTTTTCT	ATGTTGGCTT	CTCTGCAGAG	AACCACTGTA	AGCTACAAC	TAACTTTTGT	TGGACAAAT	3640
TTTCCAAACC	GCCCTTTTGC	CCTAGTGGCA	GAGACAATTC	ACAAACACAG	CCCTTAAAAA	AGGCTTAGGG	3710
ATCACTAAGG	GGATTCTAG	AAGAGCGACC	TGTAATCCTA	AGTATTTACA	AGACGAGGCT	AACCTCCAGC	3780
GAGCGTGACA	GCCAGGGGAG	GGTGCGAGGC	CTGTTCAAAT	GCTAGCTCCA	TAAATAAAGC	AATTTCCCTC	3850
GGCAGTTTCT	GAAAGTAGGA	AAGGTTACAT	TTAAGGTTGC	GTTTGTAGC	ATTTCAGTGT	TTGCCGACCT	3920
CAGCTACAGC	ATCCCTGCAA	GGCCTCGGGA	GACCCAGAAG	TTTCTCGCCC	CCTTAGATCC	AAACTTGAGC	3990
AACCCGGAGT	CTGGATTCTT	GGGAAGTCTT	CAGCTGTCTT	GCGGTTGTGC	CGGGGCCCCA	GGTCTGGAGG	4060
GGACCACTGG	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAGT	CGGGCCTCCT	AGCTCTGCAG	TCCGAGGCTT	4130
GGAGCCAGGT	GCCTGGAGCC	CGAGGCTGCC	CTCCACCCTG	TGCGGGCGGG	ATGTGACCAG	ATGTTGGCCT	4200
CATCTGCCAG	ACAGAGTGCC	GGGGCCACGG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGCGCG	4270
GCCAGCAGGA	GCGCCTGGCT	CCATTTCCTA	CCCTTTCTCG	ACGGGACCGC	CCCGGTGGGT	GATTAAACAG	4340
TTTGGGGTGG	TTTGCTCATG	GTGGGGACCC	CTCGCGCCTT	GAGAACCTGC	AAAGAGAAAT	GACGGGCTTG	4410
TGTCAAGGAG	CCCAAGTCGC	GGGGAAGTGT	TGCAGGGAGG	CACCTCCGGG	GGTCCCGCGT	GCCCGTCCAG	4480
GGAGCAATGC	GTCTCTGGGT	TGCTCCCCAG	CCGCGCTAC	GCGCCTCCGT	CCTCCCTTTC	ACGTCCGGCA	4550
TTGTGTGTGC	CCGGAGCCCG	ACGCCCCCGG	TCCGGAACCTG	GAGGCAGCCC	TGGGTCTCCG	GATCAGGCCA	4620
GCGGCCAAAG	GGTCCCGGCA	CGCACCTGTT	CCCAGGGCCT	CCACATCATG	GCCCTTCCCT	CGGGTTACCC	4690

ERSATZBLATT (REGEL 26)

Fig. 4 (Fortsetzung)

```
CACAGCCTAG GCCGATTCCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA 4760
GCGGCGCGCG GCGGGGGAAG CGCGGGCCAG ACCCCCGGGT CCGCCCCGAG CAGCTGCGCT GTCGGGGCCA 4830
GGCCGGGCTC CCAGTGGATT CGCGGGCACA GACGCCAGG ACCGCGCTCC CCACGTGGCG GAGGGACTGG 4900
GGACCCGGGC ACCCGTCCTG CCCCTTCACC TTCCAGCTCC GCCTCCTCCG CGCGGACCCC GCCCCGTCCC 4970
GACCCCTCCC GGGTCCCCCG CCCAGCCCCC TCGGGGCCCT CCCAGCCCCCT CCCCTTCCTT TCCGCGGCCC 5040
CGCCCTCTCC TCGCGGCGCG AGTTTCAGGC AGCGCTGCGT CCGCTGCGC ACGTGGGAAG CCCTGGCCCC 5110
GGCCACCCCC GCGATG 5126
```


Fig. 5

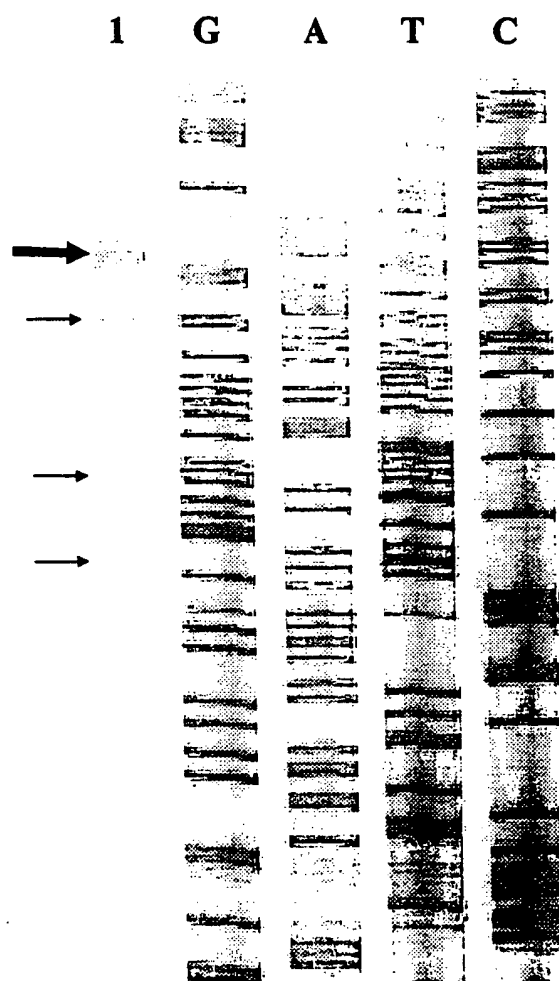


Fig. 6

```

GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG CGATGCCCGG 70
CGCTCCCCGC TGCCGAGCCG TGCGCTCCCT GCTGCGCAGC CACTACCCCG AGGTGCTGCC GCTGGCCACG 140
TTGCTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCGGGCGGCT TTCCGCGCGC 210
TGGTGGCCCA GTGCTGGTGG TGCTGCGCCT GGGACGCAGG GCGGCCCCCC GCGGCCCCCT CCTTCCGCCA 280
GGTGTCTCTG CTGAAGGAGC TGGTGGCCCG AGTGCTGCAG AGGCTGTGCG AGCGCGGCGC GAAGAACGTG 350
CTGGCCCTTC GCTTCGCGCT GCTGGACGGG GCGGCGGGG GCGCCCCCGA GGCCTTCACC ACCAGCGTGC 420
GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG GAGCGGGGCG TGGGGGCTGC TGCTGCGCGG 490
CGTGGGGGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC GCGCTCTTTG TGCTGGTGGC TCCCAGCTGC 560
GCCTACCAGG TGTGCGGGCC GCGCGTGTAC CAGCTCGGCG CTGCCACTCA GCGCGGCGCC CCGCCACACG 630
CTAGTGGACC CCGAAGCGGT CTGGGATGCG AACGGGCTTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC 700
CTTGGGCGTG CCAGCCCCGG GTGCGAGGAG GCGGCGGGG AGTGCCAGCC GAAGTCTGCG GTTGCCCAAG 770
AGGCCCGAGC GTGGCGCTGC CCCTGAGCCG GAGCGGACGC CCGTTGGGCA GGGGTCTGCG GCGCACCCGG 840
GCAGGACCGG TGGACCGAGT GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCCGCG AAGAAGCCAC 910
CTCTTTGGAG GGTGCGCTCT CTGGCACGCG CCACTCCAC CATATCCGTG GCGGCCAGCA CCACGCGGGC 980
CCCCCATCCA CATCGCGGCC ACCACGTCCC TGGGACACGC CTTGTCCCCC GGTGTACGCC GAGACCAAGC 1050
ACTTCTCTA CTCCTCAGGC GACAAGGAGC AGCTGCGGCC CTCCTTCTA CTCAGCTCTC TGAGGCCAG 1120
CCTGACTGGC GCTCGGAGGC TCGTGAGAC CATCTTCTG GGTTCACGGC CCTGGATGCC AGGGACTCCC 1190
CCGAGGTTC CCGCGCTGCC CCAGCGCTAC TGGCAAATGC GCGCCCTGTT TCTGGAGCTG CTTGGGAACC 1260
ACGCGCAGTG CCCCTACGGG GTGCTCTCA AGACGCACTG CCGCTGCGA GCTGCGGTCA CCCCAGCAGC 1330
CGGTGTCTGT CCGCGGGAGA AGCCCCAGGG CTCTGTGGCG GCGCCCGAGG AGGAGGACAC AGACCCCGCT 1400
CGCCTGGTGC AGTGCTCCG CCAGCACAGC AGCCCCTGGC AGGTGTACGG CTTGCTGCGG GCCTGCCTGC 1470
GCGGCTGGT GCCCCAGGC CTCTGGGCT CCAGGCACAA CGAACGCGCG TTCTCAGCA ACACCAAGAA 1540
GTTTCTCTC CTGGGGAAGC ATGCCAAGCT CTCGCTGCAG GAGCTGACGT GGAAGATGAG CGTGCGGGAC 1610
TGCGCTTGGC TGCGCAGGAG CCCAGGGGTT GCGTGTGTT CCGCCCGAGA GCACCGTCTG CGTGAGGAGA 1680
TCCTGGCCAA GTTCTGACG TGGCTGATGA GTGTGTACGT CGTGAGCTG CTCAGGTCTT TCTTTATGT 1750
CACGGAGACC ACGTTTCAA AGAACAGGCT CTTTCTTAC CGGAAGAGTG TCTGGAGCAA GTTGCAAGC 1820
ATTGGAATCA GACAGCACTT GAAGAGGGTG CAGCTGCGGG AGCTGTGCGA AGCAGAGGTC AGGCAGCATC 1890
GGGAAGCCAG GCGCGCCCTG CTGACGTCCA GACTCCGCTT CATCCCCAAG CCTGACGGGC TCGGCGCGAT 1960
TGTGACATG GACTACGTGC TGGGAGCCAG AACGTTCCG ACAGAAAGA GGGCCGAGCG GTCTACCTCG 2030
AGGGTGAAG CACTGTTTCA CGTCTCAAC TACGAGCGGG CCGCGCGCCC CCGCTCCTG GCGCGCTCTG 2100
TGCTGGGCTT GGACGATATC CACAGGGCCT GCGCGACCTT CGTGCTGCGT GTGCGGGGCC AGGACCCGCC 2170
GCGTGAAGTG TACTTTGTC AGGTGGATG GACGGGCGCG TACGACACCA TCCCCAGGA CAGGCTCACG 2240
GAGGTGATCG CCAGCATCAT CAACCCCGAG AACACGTACT CCGTGCCTCG GTATGCCGTG GTCCAGAGG 2310
CGGCCCATGG GCACGTCCCG AAGGCCTTCA AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT 2380
TCCCTGAATG AGGCCAGCAG TGGCTCTTTC GACGTCTTTC TACGCTTCAT GTGCCACCAC GCGGTGCGCA 2520
TCAGGGGCAA GTCTTACGTC CAGTGCCAGG GGATCCCGCA GGGTCCATC CTCTCCACGC TGCTCTGAG 2590
CCTGTGCTAC GCGGACATGG AGAACAGCT GTTTGCGGGG ATTCGGCGGG ACGGGCTGCT CCGCGGTTG 2660
GTGGATGAT TCTTGTGGT GACACCTCAC CTCACCCACG CGAAACCTT CCTCAGGACC CTGCTCCGAG 2730
GTGTCCCTGA GTATGGCTGC GTGGTGAAT TCGGGAAGAC AGTGTGAAC TTCCCTGTAG AAGACGAGGC 2800
CCTGGTGGC ACGGCTTTG TTCAGATGCC GCGCCACGGC CTATTCCTT GGTGCGGCT GCTGCTGGAT 2870
ACCCGAGCCC TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCGGACCTC CATCAGAGCC AGTCTCACCT 2940
TCAACCGCGG CTTCAAGGCT GGGAGGAACA TCGCTCGCAA ACTCTTTGGG GTCTTGGCGG TGAAGTGTCA 3010
CAGCCTGTTT CTGGATTTC AGGTGAACAG CCTCCAGACG GTGTGCACCA ACATCTACAA GATCCTCTG 3080
CTGCAGGCGT ACAGGTTTCA CGCATGTGTG CTGCAGCTCC CATTTCATCA GCAAGTTTGG AAGAACCCTA 3150
CATTTTCTT CCGCGTCATC TCTGACACGG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACCGAGG 3220
GATGTGCTG GGGGCCAAGG GCGCGCGCGG CCTCTGCCCC TCCGAGGCGG TGCASTGGCT GTGCCACCAA 3290
GCATTCCTGC TCAAGCTGAC TCGACACCGT GTCACCTACG TGCCACTCCT GGGGTCACTC AGGACAGCCC 3360
AGACCGAGCT GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCCTGGAG GCGCGAGCCA ACCCGGCACT 3430
GCCCTCAGAC TTCAGACCA TCCTGGACTG ATGGCCACCC GCGCCAGACC AGGCGGAGAG CAGACACCAG 3500
CAGCCCTGTC ACCCGGGCT CTACGTCCCA GGGAGGGAGG GCGGCGCCAC ACCCAGGCCC GCACCGCTGG 3570
GAGTCTGAGG CTTGAGTGAG TGTTTGGCCG AGGCTGCTAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3640
GCCTGAGCGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCGCTC TTCATTCCC CACAGGCTGG 3710
CGCTCGGCTC CACCCAGGG CCAGCTTTTC CTCACAGGA GCGCGGCTTC CACTCCCCAC ATAGGAATAG 3780
TCCATCCCCA GATTGCCCC TGTTCACCCC TCGCCCTGCC CTCCCTTGGC TTCCACCCCC ACCATCCAGG 3850
TGGAGACCCT GAGAAGGACC CTGGGAGCTC TGGGAATTTG GAGTGACCAA AGGTGTGCCC TGTACACAGG 3920
CGAGGACCCT GCACCTGGAT GGGGCTCCCT GTGGGTCAA TTGGGGGGAG GTGCTGTGGG AGTAAATATC 3990
TGAATATATG AGTTTTTCAG TTTTGAAAAA AAAAAAAAAA AAAAAAAAAA AA 4042

```

Fig. 7

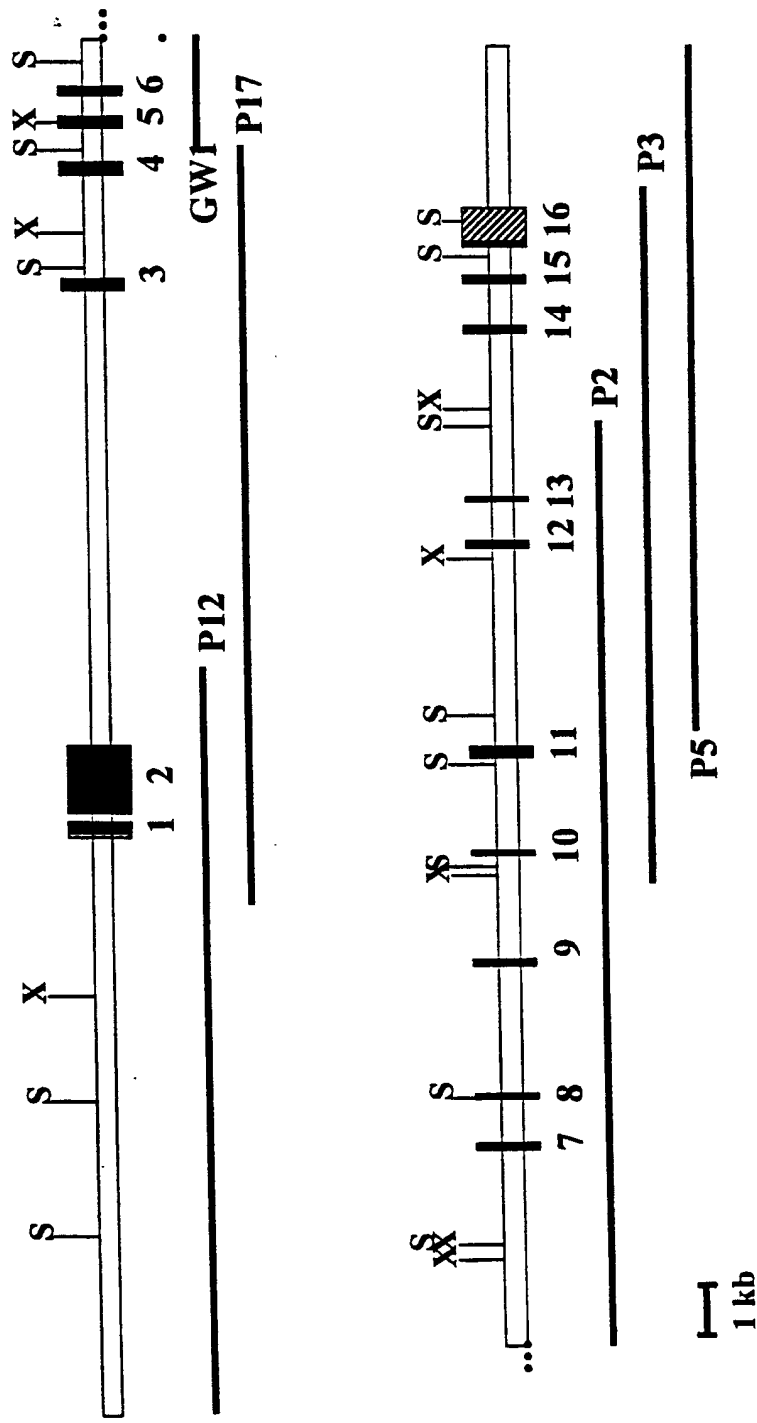


Fig. 8A

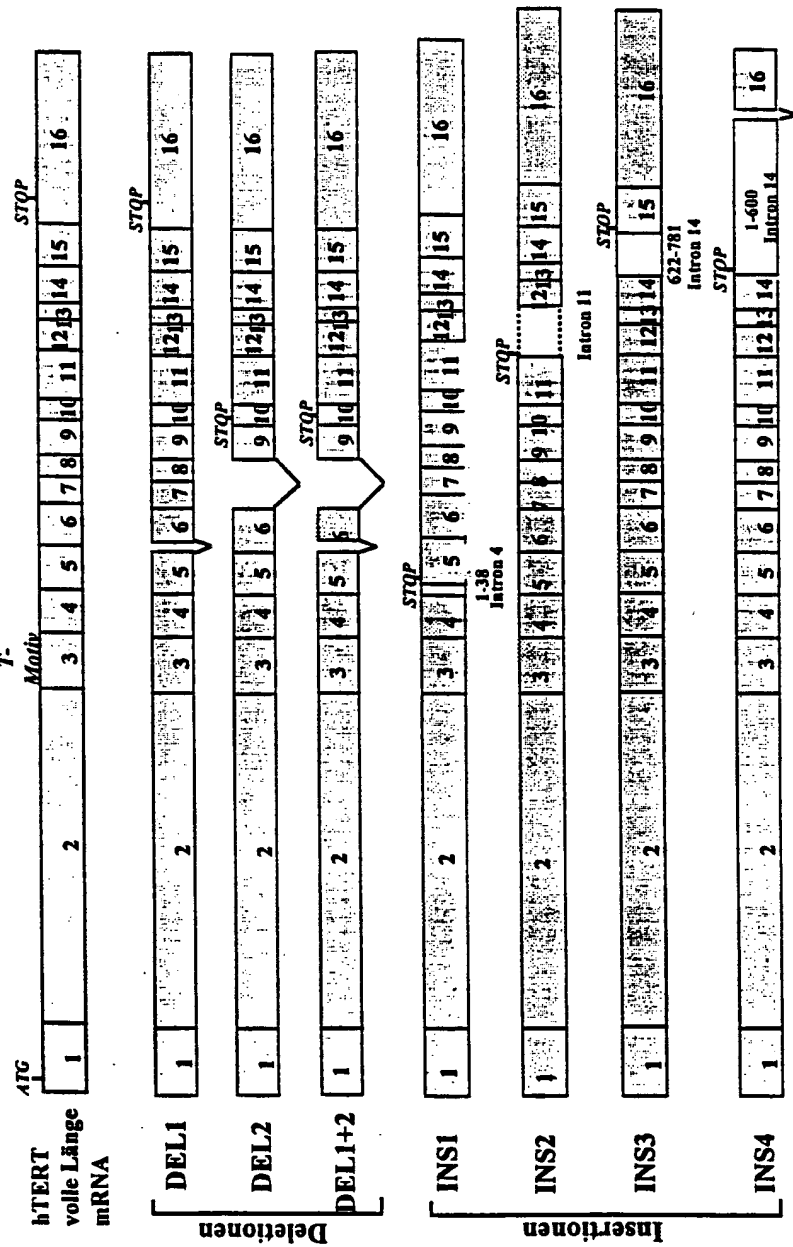


Fig. 8B

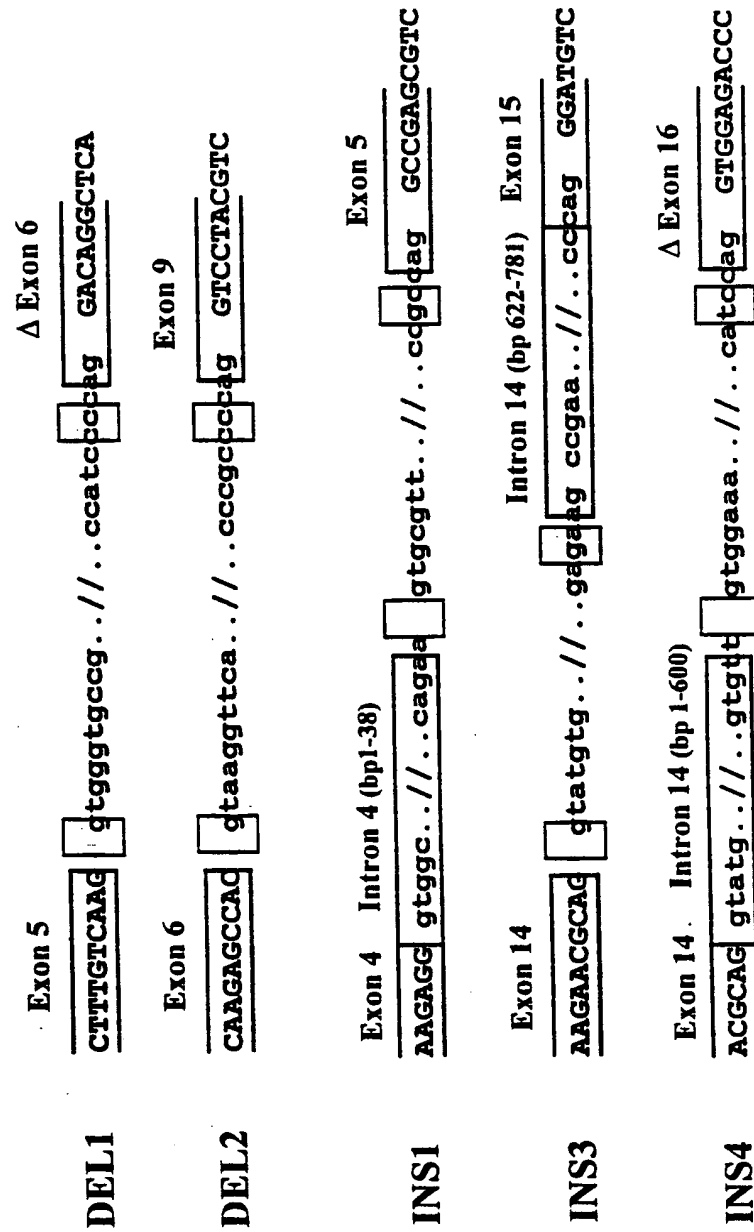


Fig. 9

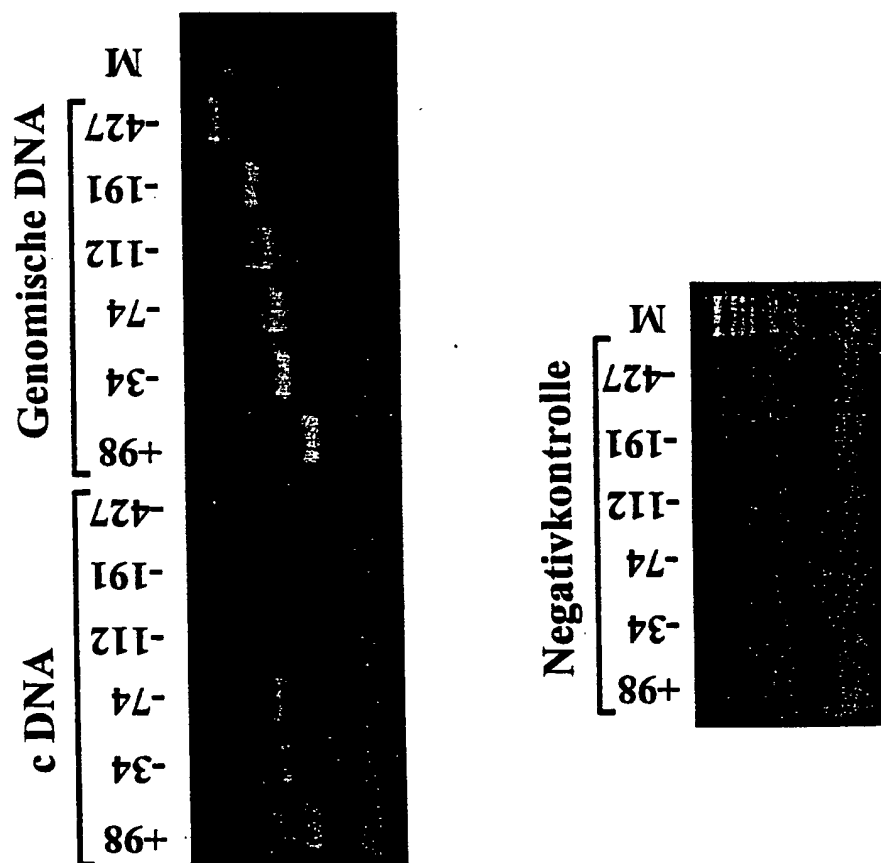


Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC ACGCCAGCCT TGGTGACAGA -11204
 ATGAGACCCT GTCTCAAAAA AAAAAAAAAA AATTGAAATA ATATAAGCA TCTTCTCTGG CCACAGTGGG -11134
 ACAAACCAG AAATCAACAA CAAGAGGAAT TTTGAAAACT ATACAAACAC ATGAAAAATTA AACAAATATAC -11064
 TTCTGAATGA CCAGTGAGTC AATGAAGAAA TTAAGGAGGA AATTGAAAAA TTTATTTTAAG CAAATGATAA -10994
 CGGAAACATA ACCTCTCAAA ACCCAGCGTA TACAGCAAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924
 AGCAGGTACA TCAAAAAAGT AGAAAAGCCA GGCAGCTGG CTCATGCCTG TAATCCCAGC ACTTTGGGAG -10854
 GCCAAGGCGG GCAGATCGCC TGAGGTGAGC AGTTCGAGAC CAGCCTGACC AACACAGAGA AACCTTGTG -10784
 CTAATAAAAA TACAAAATTA GCTGGGCATG GTGGCACATG CCTGTAATCC CAGCTACTCG GGAGGCTGAG -10714
 GCAGGATAAC CGCTTGAACC CAGGAGGTGG AGGTTGCGGT GAGCCGGGAT TGCGCCATTG GACTCCAGCC -10644
 TGGGTAACAA GAGTGAAACC CTGTCTCAAG AAAAAAAAAA AAGTAGAAAA ACTTAAAAAT ACAACCTAAT -10574
 GATGCACCTT AAAGAAGTAG AAAAGCAAGA GCAAACTAAA CCTAAAATTG GTAAAAGAAA AGAATAATA -10504
 AGATCAGAG CAGAAATAAA TGAAGCTGAA AGATAACAAT ACAAAGATC AACAAAATTA AAAGTTGGTT -10434
 TTTTGAAGAG ATAAACAAAA TTGACAAACC TTTGCCAGA CTAAGAAAAA AGGAAAGAGG ACCTAAATAA -10364
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAATT CAAAGGATCA CTAGAGGCTA -10294
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTCTAGA TGCATACAAC -10224
 CTACCAAGAT TGAACCATGA AGAAATCCAA AGCCCAACAA GACCAATAAC AATAATGGGA TTAAGCCAT -10154
 AATAAAAGT CTCCTAGCAA AGAGAAGCCC AGGACCCAT GGCTTCCCTG CTGGATTTTA CCAATCTATT -10084
 AAAGAAGAT GAATTCCAAT CCTACTCAA CTATTCTGAA AATAGAGGA AAGAATACTT CCAACTCAT -10014
 TCTACATGGC CAGTATTACC CTGATTCCAA AACCAGACAA AAACACATCA AAAACAAACA AAAAAAAA -9944
 CAGAAAGAAA GAAACTACA GGCCAATATC CTGTATGAAT ACTGATACAA AATCCTCAA CAAAACACTA -9874
 GCAAAACAAA TTAACAACA CCTTCGAAG ATCATTCTT GTGATCAAGT GGGATTTATT CCAGGGATGG -9804
 AAGGATGGT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACAAACACTA -9734
 TATGATTATT TCACTTTATG CAGAAAAAGC ATTTGATAAA ATTCTGCACC CTTCTATGATA AAAACCTCA -9664
 AAAAACCCAGG TATACAGAA ACATACAGGC CAGGCACAGT GGCTCACACC TGCATCCCA GCACCTCTGGG -9594
 AGGCCAAGGT GGGATGATTG CTTGGGCCCA GGAGTTTGG ACTAGCCTGG GCAACAAAT GAGACCTGGT -9524
 CTACAAAAA CTTTTTAAA AAATTAGCCA GGCATGATG CATATGCCTG TAGTCCCAGC TAGTCTGGAG -9454
 CTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGG TGCAGTGAGC CATGAACATG TCACTGTACT -9384
 CCAGCCTAGA CAACAGACA AGACCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAAGTGG AAGGGGAAGG GGAAGGGAAA GAGGAAGAAG -9244
 AAGAACAATA TTCAACATA ATAAAAGCCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAAGCTGAA -9174
 AGCCTTTCTT CTAAGATCTG GAAAATGACA AGGGCCCACT TTCACCACTG TGATTCAACA TAGTACTAGA -9104
 AGTCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAA GGCAATCCAA CTGGAAGAGA AGAAGTCAAA -9034
 TTATCCTGTT TGCAGATGAT ATGATCTTAT ATCTGAAAA GACTTAAGAC ACCACTAAA AACTATTAGA -8964
 GCTGAAATTT GGTACAGCAG GATACAAAA CAATGTACAA AATCAGTAG TATTTCTATA TTCCAACAGC -8894
 AAACAATCTG AAAAAAGAAC CAAAAAGCA GCTACAAATA AATTAACA GCTAGGAATT AACCAAGAA -8824
 GTGAAGATCT TCTACAATGA AAATATAAA ATGTTGATAA AAGAAATGA AGAGGGCACA AAAAAAGAA -8754
 AGATATTCCA TGTTCATAGA TTGGAAGATT AAATCTGTT AAAATGTCCA TACTACCAA AGCAATTAC -8684
 AAATCAATG CAATCCCTAT TAAATACTA ATGACGTTCT TCACAGAAAT AGAAGAAACA ATTCTAAGAT -8614
 TTGTACAGAA CCACAAAAGA CCCAGATAG CCAAAGCTAT CCTGACCAA AAGAACAAA CTGGAAGCAT -8544
 CACATTAACCT GACTTCAAT TATATACAA AGCTATAGTA ACCCAACTA CATGGTACTG GCATAAAAAAC -8474
 AGATGAGACA TGGACCAGAG GAACAGAATA GAGAATCCAG AAACAAATCC ATGCATCTAC AGTGAATCA -8404
 TTTTGTACAA AGGTGCCAAG AACATACCTT GGGGAAAGA TAATCTCTTC AATAAATGGT GCTGGAGGAA -8334
 CTGGATATCC ATATGCAAAA TAACAATACT AGAATCTGT CTCTCACCAT ATACAAAAGC AATCAAAAT -8264
 GGATGAAAGG CTTAAATCTA AAACCTCAAA CTTTGCAACT ACTAAAAGAA AACACCGGAG AAATCTCTCA -8194
 GGACATTGGA GTGGGCAAG ACTTCTTGG TAATTCCTCG CAGGCACAGG CAACCAAGC AAAACAGAC -8124
 AAATGGGATC ATATCAAGTT AAAAGCTTC TGCCAGCAA AGGAAACAA CAACAAAGAG AAGAGACAAC -8054
 CCACAGAATG GGAGAATATA TTTGCAAACT ATTCATCTAA CAAGGAATTA ATAACCAGTA TATATAAGGA -7984
 GCTCAAACTA CTCTATAGA AAAACACCTA ATAAGCTGAT TTTCAAAAT AAGCAAAAGA TCTGGGTAGA -7914
 CATTTCTCAA AATAAGTCAT ACAAATGGCA AACAGGCATC TGAAAATGTG CTCAACACCA CTGATCATCA -7844
 GAGAAATGCA AATCAAACT ACTATGAGAG ATCATCTAT CCCAGTTAAA ATGGCTTTTA TTCAAAAGAC -7774
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGGAC ACTGTTGGTG GGAATGGAAA -7704
 TTGCTACCAC TATGGAGAAC AGTTTGAAAG TTCTCAAAA AACTAAAAAT AAAGCTACCA TCAGCAATC -7634
 CCATTGCTAG GTATATACTC CAAAAAGGG AATCAGTGT TCAACAAGCT ATCTCCACTC CCACATTTAC -7564
 TGCAGCACTG TTCATAGCAG CCAAGGTTTG GAAGCAACCT CAGTGTCCAT CAACAGACGA ATGGAAGAA -7494
 AAAATGTGGT GCACATACAC AATGGAGTAC TACGCAGCCA TAAAAAGAA TGAGATCCTG TCAGTTGCAA -7424
 CAGCATGGGG GGCAGTGGTC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354
 CTCCTTACT TGTGGGAGCA AAAATTAATA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284
 GGTGGGGGAC AGGGTGACTA GAGTCAACAA TAATTTATTG TATGTTTTAA AATAACTAAA AGAGTATAAT -7214
 TGGGTTGTTT GTAACACAAA GAAAGGATAA ATGCTTGAAG GTGACAGATA CCCCATTTAC CCTGATGTGA -7144
 TTATTACACA TTGTATGCCT GTATCAAAAT ATCTCATGTA TGCTATAGAT ATAAACCTTA CTATATTAAA -7074
 AATTAATTT TTAATGGCCA GGCACGGTGG CTCATGTCCG TAATCCCAGC ACTTTGGGAG GCCGAGGCGG -7004
 GTGGATCACC TGAGGTGAGG AGTTTGAAAC CAGTCTGGCC ACCATGATGA AACCCTGTCT CTACTAAAG -6934
 TACAAAAAT AGCCAGGCGT GGTGGCAGAT ACCTGTAGTC CCAACTACTC AGGAGGCTGA GACAGGAGAA -6864
 TTGCTTGAAC CTGGGAGGCG GAGGTTGCAG TGAGCCGAGA TCATGCCACT GCACTGCAGC CTGGGTGACA -6794
 GAGCAAGACT CCATCTCAA ACAAAAACAA AAAAAAGAG ATTAATAATG TAATTTTATG TACCGTATA -6724
 AATATATCT CTACTATATT AGAAGTTAAA AATTAACAA ATTATAAAG GTAATTAACC ACTTAATCTA -6654
 AAATAAGAAC AATGTATGTG GGGTTTCTAG CTTCTGAAGA AGTAAAAGT ATGGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTTTAAAC -6514
 AAAGACAGGC TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAAGG -6444
 TAATCTCTAT TAATTACCAA TAATTACAGA TATCTCTAAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374
 CACCGTCCTC TCATTACCGG TGCTTTTTTT CTTGTGTGCT TGGAGATTTT CGATTGTGTG TTCGTGTTTG -6304
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAAATGGT GGTGATTTCC TCCAGAAGAA TTAGAGTACC -6234
 TGGCAGGAAG CAGGTGGCTC TGTGGACCTG AGCCACTTCA ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164
 TGCAAGGCAG AGGCCTGATG ACCCGAGGAC AGGAAAGCTC GGATGGGAAG GGGCGATGAG AAGCCTGCCT -6094
 CGTTGGTGAG CAGCGCATGA AGTGCCCTTA TTTACGCTTT GCAAAGATTG CTCTGGATAC CATCTGGAAA -6024
 AGGCGGCCAG CGGGAATGCA AGGAGTCAGA AGCCTCCTCG TCAAACCCAG GCCAGCAGCT ATGGCGCCCA -5954
 CCCGGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884
 AGCAGTGACC AAGGTGTATT CTGAGGGAAG CTTGAGTTAG GTGCCCTCTT TAAACAGAA AGTCATGGAA -5814
 GCACCTTTCT CAAGGGAAAA CCAGACGCCC GCTCTGCGGT CATTTACCTC TTCTCTCTCT CCTCTCTTG -5744
 CCCTCGCGGT TTCTGATCGG GACAGAGTGA CCCCGTGGA GCTTCTCCGA GCCCGTGTG AGGACCTCT -5674
 TGCAAGGGC TCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGCCTGG CTTAATAACA AACTGGGATG -5604
 TGGCTGGGG CGGACAGCGA CGGCGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCCACAT -5534
 CCGAATGGAT TTGGATTTTA TCTTAATATT TTCTTAAAT TCATCAAATA ACATTAGGA CTGCAGAAAT -5464
 CCAAAGGCGT AAAACAGGAA CTGAGCTATG TTGCGCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394
 ATTTTTCGCC CTAAGTACTT TTTATTGTTT TTCATAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324
 AGAGGCTCTG GCGGCAGGCG TATGAGCAGC GCAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254
 ACAGCAGGAC CACTGACCGT CCTCCCTGGG AGCTGCCACA TTGGGCAACG CGAAGGGCGG CACGCTGCTG -5184
 GTGACTCAGG ACCCCATACC GGCTTCTTGG GCCCACCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114
 CCGTGGAAAC GAACATGACC CTTGCCCTGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA TGTGTTGACA -5044
 GGAATGGCC ATGTAAATTA CACGACTCTG CTGATGGGGA CCGTTCCTTC CATCATATT CATCTTCC -4974
 CCCAAGGACT GAATGATTCC AGCAACTTCT TCGGGTGTGA CAAAGCATGA CAAAACTCAG TACAACACCC -4904
 ACTCTTTTAC TAGGCCACCA GAGCAGGSC CACACCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834
 GCTTTCAGCC ACCAGGCTGG GGTGACAACA CGGGCTGAAC AGTCTGTTCC TCTAGACTAG TAGACCCTGG -4764
 CAGGCACTCC CCCAGATTCT AGGGCTTGGT TGCTGCTTCC CGAGGGCGCC ATCTGCCCTG GAGACTCAGC -4694
 CTGGGGTGCC AACTGAGGC CAGCCCTGTC TCCACACCCT CCGCCTCCAG GCCTCAGCTT CTCCAGCAGC -4624
 TTCTTAAAC CTGGGTGGGC CGTGTTCAG CGCTACTGTC TCACCTGTCC CACTGTGTCT TGTCTCAGG -4554
 ACCTAGCTCG CACGGTTCTT CTTCACATGG GGTGTCTGTC TCCTTCCCA ACACCTCACAT CGCTTGAAGG -4484
 GAGGAGATTG TGCGCCTCCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGCCCCC GATGCAGGTT -4414
 CCTGGCTGCC GGCTGCACGC TGACCTCCAT TTCAGGCGC TCCTCCGCTC CTGTCTCTG CCGGGGCTG -4344
 CCGGTGTTT CTTCTGTTT TGTGCTCCT TCCACGTCCA GCTGCGTGTG TCTCTGCCCC CTAGGGTCTC -4274
 GGGGTTTTTA TAGGCATAGG ACGGGGGCGT GGTGGGCGAG GCGCTCTTG GGAATGCAA CATTTGGGTG -4204
 TGAAAGTAGG AGTGCTGTC CTCACCTAGG TCCACGGGCA CAGGCTGGG GATGGAGCCC CCGCCAGGGA -4134
 CCGGCCCTTC TCTGCCAGC ACTTTCCTGC CCCCTCCCT CTGGAACACA GAGTGGCAGT TTCCACAAGC -4064
 ACTAAGCATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CTGGACATTT CCGCCACAGT CTTGGGAATT -3994

c-Myc
 CACGTGACTA CGCAGCATCAT GTACACACTC CCGTCCACGA CCGACCCCG CTGTTTTATT TTAATAGCTA -3924
 CAAAGCAGGG AAATCCCTGC TAAATGTGCC TTTAAACAAAC TGGTTAAACA AACGGGTCCA TCCGCACGGT -3854
 GGACAGTTCC TCACAGTGAA GAGGAACATG CCGTTTATAA AGCCTGCAGG CATCTCAAGG GAATTACGCT -3784
 GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAG AAAGAATTTC ACCCCATGGC -3714
 AGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCGGCA GCTGGGGGCT ACTGCAGGCA CTTTACTA -3644
 AAGCCAGTTT CTTGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGGAG TGGGGATGGG -3574
 GGAACCCGGA GGCTGTGCCA TCTTTGCCAT GCCCGAGTGT CCTGGGAGG ATAATGCTCT AGAGATGCC -3504
 ACGTCTGAT TCCCCAAAC CTGTGGACAG AACCCGCGCG GCCCAGGGC CTTTGCAGT GTGATCTCCG -3434
 TGAGGACCTT GAGGTCTGGG ATCCTTCGGG ACTACCTGCA GGCCCGAAAA GTAATCCAGG GGTTCGGGA -3364
 AGAGGCGGGC AGGAGGGTCA GAGGGGGGCA GCCTCAGGAC GATGGAGGCA GTCAGTCTGA GGCTGAAAG -3294
 GGAGGGAGGG CCTCGAGCCC AGGCCTGCAA GCGCCTCCAG AAGCTGGAAG AAGCGGGGAA GGGACCTCC -3224
 ACGGAGCCTG CAGCAGGAAG GCACGGCTGG CCTTAGCCC ACCAGGGCCC ATCGTGACC TCCGGCCTCC -3154
 GTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTGTGGGG ATTTGCAGAA GCAACAGGAA -3084
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTTACAGAAA CATCCAAGGA CAGGGGTGAA -3014
 GTGCCTCCGG GCAAGGGCAG GGCAGGCAGC AGTGATTTTA TTTAGCTATT TTATTTTATT TACTTACTTT -2944
 CTGAGACAGA GTTATGCTCT TGTGGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874
 CGTCTCCTGG GTTCAAGCAA TTCTCGTGCC TCAGCCTCCC AAGTAGCTGG GATTTAGGC GTGCACCACC -2804
 ACACCCGGCT AATTTTGTAT TTTTAGTAGA GATGGGCTTT CACCATGTTG GTCAAGCTGA TCTCAAAATC -2734
 GTGACCTCAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664
 GGCCTATTTA ACCATTTTAA AACTTCCCTG GGCTCAAGTC ACACCCACTG GTAAGGAGTT CATGGAGTTC -2594
 AATTTCCCTT TTAAGCAGGA GTTACCCTCC TTTGATATTT TCTGTAATTC TTCGTAGACT GGGGATACAC -2524
 CGTCTCTTGA CATATTCACA GTTCTGTGTA CCACCTGTTA TCCCATGGGA CCCACTGCAG GGCAGCTGG -2454
 GAGGCTGCAG GCTTCAGGTC CAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384
 AAGTGTGGAC ACTGTCTCTA ATCTCAATG CTCAGTGTGT GCTGAAACAT GTAGAAATTA AAGTCCATCC -2314
 CTCTACTCT ACTGGGATTG AGCCCTTCC CTATCCCCC CCAGGGGCG AGGAGTTTCT CTCACTCTG -2244
 TGGAGGAAGG AATGATACTT TGTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGTTTGG -2174
 TTTGTTTTGT TTTGAGAGGC GGTTCACCTC TTGTTGCTCA GGCTGGAGGG AGTGCAATGG CGCGATCTTG -2104
 GCTTACTGCA GCCTCTGCCT CCCAGTTCA AGTGATTCTC CTGCTTCCGC CTCCCATTTG GCTGGGATTA -2034
 CAGGCACCCG CCACATGCC CAGCTAATTT TTTGATATTT TAGTAGAGAC GGGGGTGGT GGGGTTCACC -1964

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGTATG TCCACCTGCC TCTGCCTCCT AAAGTGCTGG -1894
GATTACAGGT GTGAGCCACC ATGCCACGCT CAGAATTAC TCTGTTTAGA AACATCTGGG TCTGAGGTAG -1824

CAAT-Box

GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTATGCCAA TGATAGAATT TTTTATTGT TGTTAGAACA -1754

CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTTCAAAG ACACACTAAC TGCACCCATA -1684

ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCCGGGAGG CGTTTCCTCG CCATGCACAT -1614

GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTAAAT TGTGTTTCT -1544

ATGTTGGCTT CTCTGCAGAG AACCAGTGTA AGCTACAAC TAACTTTGT TGGAAACAAT TTTCCAAACC -1474

Spi
GCCCTTTGC CCTAGTGGA GAGACAATC ACAAACACAG CCCTTAAAA AGGCTTAGGG ATCACTAAGG -1404

GGATTCTAG AAGAGCGACC TGTAATCCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334

GCCCAGGGAG GGTGCGAGGC CTGTTCAAAT GCTAGCTCCA TAAATAAGC AATTTCTCC GGCAGTTCT -1264

GAAAGTAGGA AAGGTTACAT TTAAGGTTGC GTTTGTTAGC ATTTCACTGT TTGCCGACCT CAGCTACAGC -1194

ATCCCTGCAA GGCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACCTGAGC AACCCGGAGT -1124

CTGGATTCTT GGAAGTCTT CAGCTGTCTT GCGGTTGTGC CGGGCCCCA GGTCTGGAGG GGACCACTGG -1054

CGGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCTT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984

GCTTGAGACC CGAGGCTGCC CTCCACCCTG TCGGGCGGG ATGTGACCAG ATGTTGGCTT CATCTGCCAG -914

ACAGAGTGCC GGGGCCAGG GTCAAGGCC TTGTGGCTGG TGTGAGGCGC CCGGTGCGCG GCCAGCAGGA -844

CCAC-Box **Spi**
GCGCCTGGCT CCATTCCCA CCCTTCTCG ACGGAGCCG CCCGTGGGT GATTAACAGA TTTGGGGTGG -774

TTTGCTCATG GTGGGGACCC CTCGCCCTT GAGAACCTGC AAAGAGAAAT GACGGGCTT TGTCAAGGAG -704

CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCGCT GCGCGTCCAG GGAGCAATGC -634

AP-2
GTCTCGGGT TCGTCCCGAC CCGGTCTAC GCGCCTCCGT CCTCCCTTC ACGTCCGGCA TTCGTGGTGC -564

CCGGAGCCCG ACGCCCCCG TCCGGACCTG GAGGACGCC TGGGTCTCG GATCAGGCCA GCGGCCAAAG -494

GGTCCCGCA CGCACTGTT CCCAGGGCTT CCACATCATG GCCCTCCCT CGGGTTACCC CACAGCCTAG -424

GCCGATTGCA CCTCTCTCG CTGGGGCCCT CGCTGGCGTC CCTGCACCCT GGGAGCGCGA GCGGCCGCCG -354

Spi
GGCGGGGAAG CGCGGCCAG ACCCCCGGT CCGCGGGAG CAGCTGCGCT GTCGGGGCCA GGCCGGGCTC -284

CCAGTGGATT CGCGGGCACA GACGCCAGG ACCCGCTCC CACGTGCG GAGGGACTGG GGACCCGGGC -214

ACCGTCTCG CCCCTTACC TTCCAGCTCC GCCTCTCCG CGCGGACCC GCCCGTCCC GACCCCTCCC -144

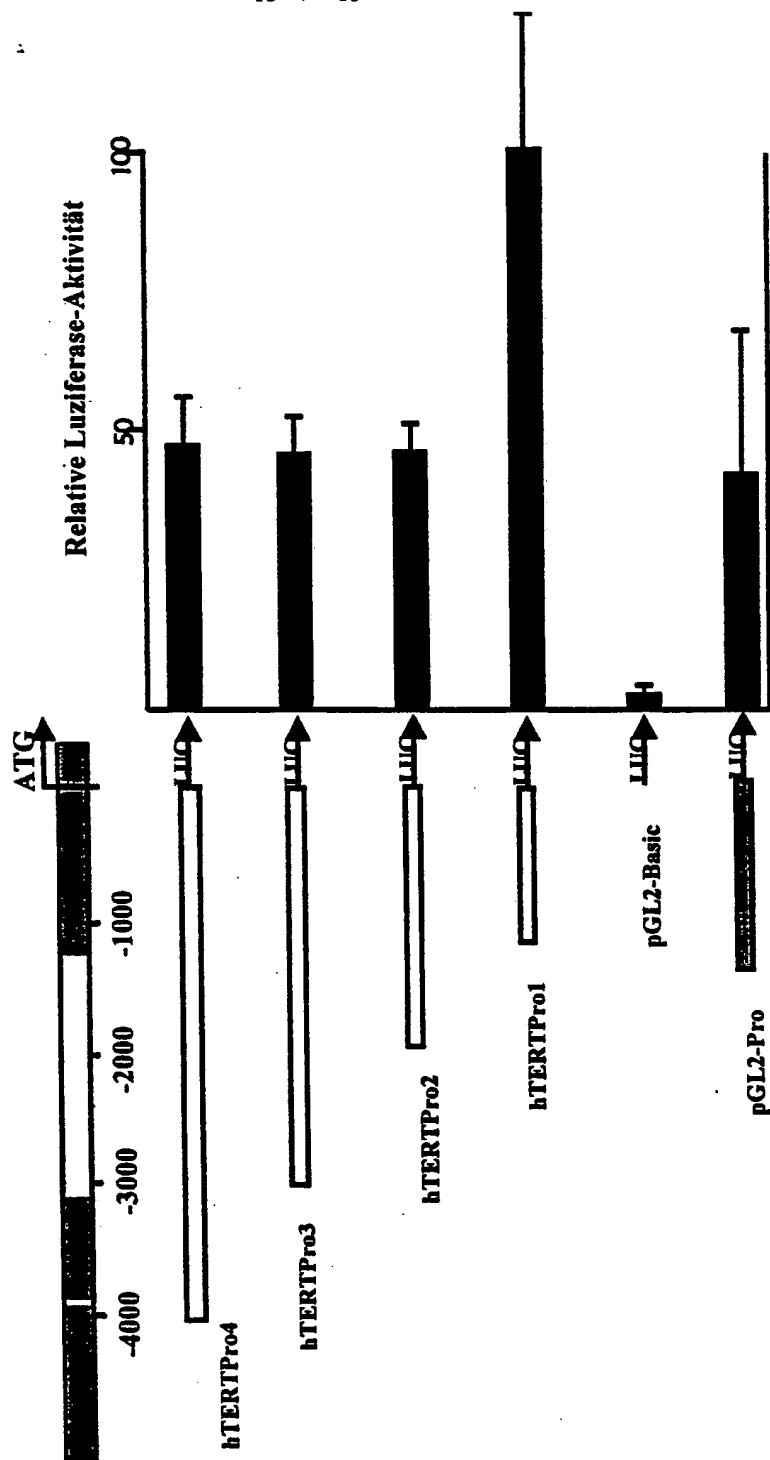
GGTCCCCGG CCCAGCCCC TCCGGGCCCT CCCAGCCCT CCCCTTCTT TCCCGGCCG CGCCCTCTCC -74

c-Myc
TCGCGGCGCG AGTTTCAGG AGCGCTGCGT CTGCTGCG CACGTGGAAG CCCTGGCCCC GGCCACCCCC -4

GCGATG

3

Fig.: 11



SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen
der humanen katalytischen Telomerase-Untereinheit und
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

<140>
<141>

15 <160> 20
<170> PatentIn Vers. 2.0

20 <210> 1
<211> 5126
<212> DNA
<213> Homo sapiens

<400> 1
gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60
25 gggtaaatgaa gtgggtgtgca ggaaatggcc atgtaaaatta cactgactctg ctgatgggga 120
ccgttccttc catcattatt catcttcacc cccaaggact gaatgattcc agcaacttct 180
tcgggtgtga caagccatga caaaactcag tacaacacacc actcttttac tagggccaca 240
gagcacgggc cacacccctg atataattaag agtcaggag agatgaggct gctttcagcc 300
accaggctgg ggtgacaaca gcggctgaac agtctgttcc tctagactag tagaccctgg 360
30 caggcactcc cccaaattct agggcctggt tgctgttcc cgaggggccc atctgccctg 420
gagactcagc ctgggtgtgcc acactgaggc cagccctgtc tccacacccct ccgcctccag 480
gcctcagctt ctccagcagc ttcctaaacc ctgggtgggc cgtgttccag cgctactgtc 540
tcacctgtcc cactgtgtct tgtctcagcg acgtagctcg cactgttccct cctcacatgg 600
ggtgtctgtc tcttccccca acactcacat gcgttgaagg gaggagattc tgcgcctccc 660
35 agactggctc ctctgagcct gaacctggct cgtgggcccc gatgcagggt cctggcgctc 720
ggctgcacgc tgacctccat tcccaggcgc tcccgtctc ctgtcatctg ccggggcctg 780
ccgggtgtgtt cttctgtttc tgtgtctctt tccacgtcca gctgcgtgtg tctctgcccg 840
ctaggtgtctc ggggttttta taggcatagg acgggggctg ggtgggcccag ggcgctcttg 900
40 ggaaatgcaa catttgggtg tgaaagtagg agtgctgtc ctacactagg tccacgggca 960
caggcctggg gatggagccc ccgcccaggga cccgcccctc tctccccagc actttctctg 1020
ccccctccct ctggaacaca gaggggcagt tccacaagc actaagcact ccttcccca 1080
aagaccagc attggcaccc ctggacattt gcccacagc cctgggaatt cactgacta 1140
cgacatcat gtacacactc ccgtccacga ccgaccccg ctgttttatt ttaatagcta 1200
45 caaagcaggg aaatccctgc taaaatgtcc tttaacaaac tggttaaaca aacgggtcca 1260
tccgcacggg ggacagtccc tcacagtga gaggaaatg ccgtttataa agcctgcagg 1320
catctcaagg gaattacgct gagtcaaaac tgccacctcc atgggatacg tacgcaacat 1380
gctcaaaaag aaagaatttc accccatggc aggggagtg ttaggggggt taaggacggg 1440
ggggggcgga gctgggggct actgcacgca ccttttacta aagccagttt cctggttctg 1500
50 atggtattgg ctgagttatg ggagactaac cataggggag tggggatggg ggaacccgga 1560
ggctgtgcca tctttgccat gcccgagtgt cctgggcagg ataagtctct agagatgccc 1620
acgtcctgat tcccccaaac ctgtggacag aaccgcccgc gcccagggc ctttgagggt 1680
gtgactctcg tgaggacct gagggtctgg atccttcggg actacctgca ggcccgaaaa 1740
gtaatccagg ggttctggga agaggcggc aggggggtca gaggggggca gcctcaggac 1800
gatggaggca gtcagtctga ggctgaaaag ggaggaggag cctcgagccc aggcctgcaa 1860
55 gcgcctccag aagctggaaa agcgggggaa gggaccctcc acggagcctg cagcaggaa 1920
gcacggctgg cccttagccc accaggggccc atcgtggacc tccggcctcc gtgccatagg 1980
agggcactcg cgtgcccctt ctgcatgaa gtgtgtgggg atttgcagaa gcaacaggaa 2040
acctatgcac tgtgaatcta ggattatttc aaaacaaagg ttacagaaa catccaagga 2100
cagggtgaa gtgcctccgg gcaagggcag ggcaggcacg agtgatttta tttagctatt 2160
60 ttattttatt tacttacttt ctgagacaga gttatgctct tcttggccag gctggagtgc 2220
agcggcatga tcttggtcca ctgcaacctc cgtctcctgg gttcaagcaa ttctcgtgcc 2280
tcagcctccc aagtagctgg gatttcaggc gtgcaccacc acaccggct aattttgtat 2340
ttttagtaga gatgggcttt caccatgttg gtcaagctga tctcaaatc ctgacctcag 2400
tgatccgcc cacctcagcc tcccaaatg ctgggattac aggcattgag cactgcacct 2460

5 ggcctattta accattttta aacttccctg ggctcaagtc acaccactg gtaaggagtt 2520
 catggagttc aatttcccct ttactcagga gttaccctcc ttgtatattt tctgtaattc 2580
 ttctgtagact ggggatacac cgtctcttga catattcaca gtttctgtga ccacctgtta 2640
 tcccatggga cccactgcag gggcagctgg gaggctgcag gcttcaggtc ccagtggggt 2700
 tgccatctgc cagtagaaac ctgatgtaga atcagggcgc aagtgtggac actgtccctga 2760
 atctcaatgt ctcagtgtgt gctgaacat gtagaaatta aagtcacatcc ctctactctc 2820
 actgggattg agccccctcc ctatcccccc ccaggggcag aggagttcct ctactcctg 2880
 tggagggaagg aatgatactt tgtattttt cactgctggg actgaatcca ctgtttcatt 2940
 10 tgttggtttg tttgttttgt tttgagaggc ggtttcactc ttgttgctca ggctggaggg 3000
 agtgcaatgg cgcgatcttg gcttactgca gccctcgctc ccagggttca agtgattctc 3060
 ctgcttccgc ctcccatattg gctgggatta caggcacccg ccaccatgcc cagctaattt 3120
 tttgtatttt tagtagagac gggggtgggt ggggttcacc atgttggtcca ggctggctc 3180
 gaacttctga cctcagatga tccactgcc tctgcctcct aaagtgcctg gattacaggt 3240
 gtgagccacc atgcccagct cagaatttac tctgtttaga aacatctggg tctgaggtag 3300
 15 gaagctcacc ccaactcaagt gttgtgggtg tttaagccaa tgatagaatt tttctattgt 3360
 tgttagaaca ctcttgatgt ttacactgt gatgactaag acatcatcag cttttcaaag 3420
 acacactaac tgcaccata atactgggt gtcttctggg tatcagcaat cttcatgaa 3480
 tgccgggagg cgtttcctcg ccatgcacat ggtgttaatt actccagcat aatcttctgc 3540
 20 tctcatctct tctcttccct cttttaaaat tgtgttttct atgttggtct ctctgagag 3600
 aaccagtgtg agctacaact taacttttgt tggaaacaaat ttcccaaacc gcccttttgc 3660
 cctagtggca gagacaattc acaaacacag ccccttaaaa aggcttaggg atcactaagg 3720
 ggatttctag aagagcgacc tgaatccta agtatttaca agacgaggt aacctccagc 3780
 gagcgtgaca gcccaggag ggtgcgaggc ctgttcaaat gctagctcca taaataaagc 3840
 aatttctccc ggcagtttct gaaagtagga aaggttacat ttaagggtgc gtttgttagc 3900
 25 atttcagtgt ttgccgacct cagctacagc atccctgcaa ggcttcggga gaccagaag 3960
 tttctcgccc ccttagatcc aaacttgagc aaccggagt ctggattcct ggggaagtcct 4020
 cagctgtcct gcggttgtgc cggggcccca ggtctggagg ggaccagtgg ccgtgtggct 4080
 tctactgctg ggctggaaat cgggctcctc agctctgcag tccgaggctt ggagccaggt 4140
 30 gcctggaccc cgaggctgcc ctccaccctg tgcgggcggg atgtgaccag atgttgacct 4200
 catctgcag acagagtgcc ggggcccagg gtcaaggccg ttgtggctgg tgtgaggcgc 4260
 ccggtgcgcg gccagcagga gcgcctggct ccatttccca ccttctctcg acgggaccgc 4320
 ccggtgggtg gattaacaga tttgggttgg ttgtctcatg gtggggacc ctcgcccct 4380
 35 gagaacctgc aaagagaaat gacgggcctg tgtcaaggag cccaagtgcg ggggaagtgt 4440
 tgcaggagg cactccggga ggtcccgcgt gccctgccag ggagcaatgc gtcctcgggt 4500
 tgcgtccca cgcgtcttac gcgcctcct cctcccttc acgtccgga ttcgtgtgtc 4560
 ccggagcccg acgccccgcg tccggacctg gaggcagccc tgggtctcgg gatcaggcca 4620
 gcggccaaag ggtgcgcgca cgcacctgt cccaggcct ccacatcat gcccctccct 4680
 cgggttacc caccagctag gccgattcga cctctctcgg ctggggccct cgtggcgtc 4740
 40 ctgcaccct gggagcgcg gcgcgcgcg ggcggggaag cgcggcccag accccgggt 4800
 ccgcccggag cagctgcgct gtcggggcca ggcgggctc ccagtggatt cgcgggca 4860
 gacgcccagg accgcgctcc ccacgtggcg gagggactgg ggaccgggc acccgtcctg 4920
 ccccttcacc ttccagctcc gcctcctcgg cgcggacccc gcccgctccc gacccctccc 4980
 ggggtccccg cccagcccc tccgggacct cccagcccc ccccttctt tccggggccc 5040
 45 cgccctctcc tgcgggcgag agtttcaggc agcgtgcgct cctgctgcgc acgtgggaag 5100
 ccctggcccc ggccaccccc gcgatg 5126

<210> 2
 <211> 4042
 <212> DNA
 50 <213> Homo sapiens

<400> 2
 gtttcaggca gcgctgcgtc ctgctgcgca cgtgggaagc cctggccccg gccacccccg 60
 cgatgccgag cgctccccgc tgcgagccg tgcgctccct gctgcgcagc cactaccgag 120
 55 aggtgctgcc gctggccacg ttctgctgcg gcctggggcc ccagggtctg cggtcgtgct 180
 agcgcgggga cccggcggtt ttccgctgcg tgggtggcca gtgcctgggt tgcgtgccct 240
 tggagcgacg gccgcccccc gccgccccct ccttccgcca ggtgtcctgc ctgaaggagc 300
 tgggtggccc agtgctgcag aggtctgtgc agcgcggcgc gaagaactgt ctggccttcg 360
 60 gcttcgcgct gctggacggg gccgcggggg gccccccga ggcttccacc accagcgtgc 420
 gcagctacct gcccaacacg gtgaccgag cactgcgggg gagcggggcg tgggggctgc 480
 tgcctgcgag cgtgggagac gacgtgctgg ttcacctgct ggcacgctgc gcctctttg 540
 tgcgtggtgc tccagctgc gcctaccagg tgtgcgggccc gccgctgtac cagctcggcg 600
 ctgcccactca ggcccgcccc ccgccacacg ctagtggacc ccgaaggcgt ctgggatgag 660
 65 aacgggacct gaacatagc gtcaggagag ccgggggtccc cctgggctcg ccagccccgg 720
 gtgcgaggag gcgcgggggc agtgccagc gaagtctgct gttgcccagg agggccaggc 780

3 / 18

gtggcgctgc ccctgagccg gagcggacgc ccgttgggca ggggtcctgg gccaccgccg 840
 gcaggacgcg tggaccgagt gaccgtgggt tctgtgtggg gtacacctgc agaccgcgcg 900
 aagaagccac ctcttttgag ggtgcgctct ctggcacgcg ccactccac ccatccgtgg 960
 gccgccagca ccacgcgggc ccccatcca catcgcgcc accacgtccc tgggacagcg 1020
 5 cttgtccccc ggtgtacgc gagaccaag acttcctcta ctctcaggc gacaaggagc 1080
 agctcgggcc ctcccttcta ctacgtctc tgaggcccg cctgactggc gctcggaggc 1140
 agtgaggacac agacccccgt cgttccaggc cctggatgcc agggactccc cgaggttgc 1200
 ccgcctggc ccagcgctac tggcaaatgc ggccctgtt tctggagctg cttgggaacc 1260
 acgcgcagtg cccctacggg gtgctcctca agacgcactg cccgctgcga gctcgggtca 1320
 10 ccccgagcgc cgggtgtctgt gcccgggaga agccccagg ctctgtggcg gcccccagg 1380
 aggaggacac agacccccgt cgcctgggtg agctgctccg ccagcacagc agccccggc 1440
 aggtgtacgc cttcgtgcg gcctgcctgc gccggctggt gccccaggc ctctggggct 1500
 ccaggcaca cgaacggcgc ttcctcagga acaccaagaa gttcatctcc ctggggaagc 1560
 atgccaagct ctgcgtcgag gactgacgt ggaagatgag cgtgcgggac tgcgcttggc 1620
 15 tgcgaggag cccaggggtt ggtgtgttc cggccgcaga gcacctctg cgtgaggaga 1680
 tcttgccaa gttcctgcac tggctgatga gtgtgtactg cgtcgagctg ctgaggtctt 1740
 tcttttatgt caggagacc acgtttcaa agaacaggct cttttctac cggaaagagt 1800
 tctggagcaa gttgcaaac attggaatca gacagcactt gaagagggtg cagctcggg 1860
 agctgtcgga agcagaggtc aggcagcatc gggaaagccag gcccgccctg ctgacctcca 1920
 20 gactccgctt catccccaa cctgacgggc tgcggccgat tgtgaacatg gactacgtcg 1980
 tggggagccag aacgttccgc agagaaaaga gggccgagcg tctcacctcg agggtgagg 2040
 cactgttcag cgtgctcaac tacgagcggg cggcgccgccc cggcctcctg ggccgctctg 2100
 tgcgtggcct ggacgatata cacagggcct ggcgcacctt cgtgctcgct gtgcgggccc 2160
 aggaccgcc gcctgagctg tactttgtca aggtggatgt gacggggcgcg tacgacacca 2220
 25 tccccagga caggctcacg gaggtcatcg ccagcatcat caaacccag aacacgtact 2280
 gcgtgcgtcg gtatgccgtg gtccagaagg ccgccatgg gcacgtcccg aaggccttca 2340
 agagccacgt ctctacctg acagacctcc agccgtacat gcgacagttc gtggctcacc 2400
 tgcaggagac cagcccgctg agggatgccc tgcgtatcga gcagagctcc tccctgaatg 2460
 agggcagcag tggcctcttc gacgtcttcc tacgcttcat gtgccaccac gccgtgcga 2520
 30 tcaggggcaa gtccctacgtc cagtgcagg ggtacccgca gggctccatc ctctccacgc 2580
 tgctctcgag cctgtgtac ggcgacatgg agaacaagct gtttgcgggg attcggcggg 2640
 acgggctgct cctgcgtttg gtggatgatt tcttgttggg gacacctcac ctacccacg 2700
 cgaaaacctt cctcaggacc ctggctccgag gtgtccctga gtatggctgc gtgtgaact 2760
 tgcgggaagac agtgggtaac ttcctgttag aagacgaggc cctgggtggc agggctttg 2820
 35 ttcagatgcc gggccacggc ctatccccct ggtgcggcct gctgctggat acccgagccc 2880
 tggaggtgca gagcgactac tccagctatg cccggacctc catcagagcc agtctcact 2940
 tcaaccgcg cttcaaggct gggaggaaaca tgcgtcgcaa actctttggg gtcttgcggc 3000
 tgaagtgtca cagcctgtt ctggatttgc aggtgaacag cctccagacg gtgtgaccca 3060
 40 acatctacaa gatcctcctg ctgcaggcgt acaggtttca cgcattgtgt ctgcagctcc 3120
 catttcatca gcaagtttg aagaacccca catttttctt gcgcgtcatc tctgacacgg 3180
 cctccctctg ctactccatc ctgaaagcca agaacgcagg gatgtcgtg ggggccaagg 3240
 gcgcgcggc cctctgcgc tccgagggcg tgcagtggct gtgccacca gcatctctgc 3300
 tcaagctgac tcgacacctg gtcacctacg tgcactcct ggggtcactc aggcagagcc 3360
 agacgcagct gactcggaag ctcccgggga gcacgctgac tgccttgag gccgcagcca 3420
 45 acccggcact gccctcagac ttcaagacca tctggactg atggccacc gccacagcc 3480
 agggcgagag cagacaccag cagccctgtc acgcccgggt ctacgtccca gggaggagg 3540
 ggcggccccc acccagggcc gcaccgctgg gactctgagg cctgagtgag tgtttggccg 3600
 agggctgcat gtccggctga aggtgtgag tccggctgag gccagagcga gtgtccagcc 3660
 50 aagggctgag tgtccagcac acctgcccgtc ttcacttccc cacaggctgg cgtcgggtc 3720
 caccacagg ccagcttttc ctaccagga gcccggttc cactccccac ataggaatag 3780
 tccatcccca gattcgccat tgttcacccc tcgcccgtgc ctcttttgc ttcaccccc 3840
 accatccagg tggagacctt gagaaggacc ctgggagctc tgggaatttg gactgaccaa 3900
 aggtgtgccc tgtacacagg cgaggacctt gcacctggat ggggggtccc gtgggtcaaa 3960
 55 ttggggggag gtgctgtggg agtaaaatac tgaatatatg agtttttcag ttttgaaaaa 4020
 aaaaaaaaa aaaaaaaaa aa 4042

<210> 3
 <211> 11276
 <212> DNA
 60 <213> Homo sapiens

<400> 3
 acttgagccc aagagttcaa ggctacgggt agccatgatt gcaacaccac acgcccagcct 60
 tgggtgacaga atgagacctt gtctcaaaaa aaaaaaaaaa aattgaaata atataagca 120
 65 tcttctctgg ccacagtgga acaaaaccag aatcaacaa caagaggaat tttgaaaact 180

atacaaacac atgaaaaatta aacaatatat tttctgaatga ccagtggagtc aatgaagaaa 240
 ttaaaaaagga aattgaaaaa tttattttaag caaatgataa cggaacata acctctcaaa 300
 acccacggta tacagcaaaa gcagtgcctaa gaaggaagtt tatagctata agcagctaca 360
 tcaaaaaagt agaaaaagcca ggcgcagtgg ctcatgcctg taatcccagc actttgggag 420
 5 gccaaagcgg gcagatcgcc tgaggtcagg agttcgagac cagcctgacc aacacagaga 480
 acccttgtcg ctactaaaaa taaaaaatta gctgggcatg gtggcacatg cctgtaatcc 540
 cagctactcg ggaggctgag gcaggataac cgcttgaacc caggaggtgg aggttgcggt 600
 gagccgggat tgcgccattg gactccagcc tgggtaacaa gagtgaacc ctgtctcaag 660
 aaaaaaaaaa aagtagaaaa acttaaaaat acaacctaat gatgcacctt aaagaactag 720
 10 aaaagcaaga gcaaaactaaa cctaaaattg gtaaaagaaa agaaataata aagatcagag 780
 cagaaataaa tgaactgaa agataacaat acaaaagatc acaaaaatta aaagttgggt 840
 ttttgaagag ataaacaaaa ttgacaaacc tttgcccaga ctaagaaaaa aggaagaaag 900
 acctaaataa ataaagtcag agatgaaaaa agagacatta caactgatac cacagaatt 960
 caaaggatca ctagggtcga ctatgagcaa ctgtacacta ataaattgaa aaacctagaa 1020
 15 aaaatagata aattcctaga tgcatacaac ctaccaagat tgaacctga agaaatccaa 1080
 agcccaaaaca gaccaataac aataatggga ttaagccat aataaaaaag ctcttagcaa 1140
 agagaagccc aggacccaat ggcttccctg ctggatttta ccaatcattt aaagaagaat 1200
 gaattccaat cctactcaaa ctattctgaa aaatagagga aagaataact ccaactcat 1260
 tctacatggc cagtattacc ctgattccaa aaccagacaa aaacacatca aaaaacaaaca 1320
 20 aacaaaaaaa cagaagaaa gaaaactaca ggccaatc cctgargaat actgatacaa 1380
 aaatcctcaa caaaacata gaaaaccaa ttaaacaaaca ccttcgaaag atcattcatt 1440
 gtgatcaagt gggatttatt ccagggtagg aaggatgggt caacatagc aaatcaatca 1500
 atgtgatata tcatcccaac aaaatgaagt acaaaaacta tatgattatt tcactttatg 1560
 cagaaaaagc atttgataaa attctgcacc ctctcatgata aaacccctca aaaaaccagg 1620
 25 tatacaagaa acatacaggg caggcacagt ggctcacacc tgcgatccca gcactctggg 1680
 aggcgaaggt gggatgattg cttgggcccga ggagtttgag actagcctgg gcaacaaaat 1740
 gagacctggg ctacaaaaaa cttttttaa aaattagcca ggcatgatgg catatgcctg 1800
 tagtcccagc tagtctggag gctgaggtgg gagaatcact taagcctagg aggtcagagg 1860
 tgcagtggc catgaacatg tcactgtact ccagcctaga caacagaaca agacccact 1920
 30 gaataaagag aaggagaagg agaagggaga agggaggagg aaggaggagg gaggagaagg 1980
 agggaggtgga ggagaagtgg aaggggaagg ggaagggaaa gagggaagag aagaacata 2040
 tttcaacata ataaagccc tatatgacag accgaggtag tattatgagg aaaaactgaa 2100
 agcctttcct ctaagatctg gaaaatgaca agggcccact ttcaccactg tgattcaaca 2160
 tagtactaga agtcttagct agagcaatca gataagagaa agaaataaaa ggcattccaa 2220
 35 ctggaaagga agaagcaaaa ttatcctgtt tgcagatgat atgatcttat atctggaaaa 2280
 gacttaagac accactaaaa aactattaga gctgaaattt ggtacagcag gatacaaaat 2340
 caatgtacaa aaatcagtag tatttctata ttccaacagc aaacaactctg aaaaagaaac 2400
 caaaaaagca gctacaaata aaattaaaca gctaggaatt aaccaaagaa gtgaagatc 2460
 40 tctacaatga aaactataaa atgttgataa aagaaattga agagggcaca aaaaagaaa 2520
 agatattcca tgttcataga ttggaagaat aaatactgtt aaaatgtcca tactacccaa 2580
 agcaatttac aaattcaatg caatccctat taaaatacta atgacgttct tcacagaat 2640
 agaagaaaca attctaagat ttgtacagaa ccacaaaaga cccagaatag ccaaaagctat 2700
 cctgaccaa aagaacaaa ctggaagcat cacattacct gacttcaaat tatactacaa 2760
 45 agctatagta acccaaaacta catggtactg gcataaaaac agatgagaca tggaccagag 2820
 gaacagaata gagaatccag aaacaaatcc atgcatctac agtgaactca tttttgacaa 2880
 aggtgccaa aacatacttt ggggaaaaga taatctcttc aataaatggt gctggaggaa 2940
 ctggatatcc atatgcaaaa taacaatact agaactctgt ctctcacat atacaaaagc 3000
 aaatcaaaat ggatgaaagg cttaaatcta aaacctcaaa ctttgcaact actaaaagaa 3060
 50 aacaccggag aaactctcca ggacattgga gtgggcaaag acttcttgag taattccctg 3120
 caggcacagg caaccaaagc aaaaacagac aaatgggac atatcaagtt aaaaagcttc 3180
 tgcacagcaa aggaacaat caacaaagag aagagacaac ccacagaatg ggagaatata 3240
 tttgcaaaact attcatctaa caaggaatta ataaccagta tatataagga gctcaaaacta 3300
 ctctataaga aaaaacacta ataagctgat tttcaaaaat aagcaaaaaga tctgggtaga 3360
 catttctcaa aataagtcac acaaatggca aacaggcatc tgaatatgtg ctcaacacca 3420
 55 ctgatcatca gagaatgca aatcaaaact actatgagag atcatctcat cccagttaa 3480
 atggctttta tcaaaaagac aggcaataac aaatgccagt gaggatgtgg ataaaaggaa 3540
 acccttggac actgttggg ggaatggaaa ttgctaccac tatggagaac agtttgaaag 3600
 ttctcaaaa aactaaaaat aaagctacca tacagcaatc ccattgctag gtatatactc 3660
 caaaaaaggg aatcagtgta tcaacaagct atctccactc ccacatttac tgcagcactg 3720
 60 ttcatacgag ccaaggtttg gaagcaacct cagtgtccat caacagacga atggaaaaag 3780
 aaaatgtggt gcacatacac aatggagtac tacgcagcca taaaaaagaa tgatgcctg 3840
 tcagttgcaa cagcatgggg ggcactggtc agtatgttaa gtgaaataag ccaggcacag 3900
 aaagacaaac ttttcatgtt ctcccttact tgtgggagca aaaaataaaa caattgacat 3960
 65 agaaatagag gagaatgggt gttctagagg ggtgggggac aggggtgacta gagtcaacaa 4020
 taatttattg tatgttttaa aataactaaa agagtataat tgggttgttt gtaacacaaa 4080

gaaaggataa atgcttgaag gtgacagata ccccatTTac cctgatgtga ttattacaca 4140
 ttgtatgcct gtatcaaaat atctcatgta tgctatagat ataaacccta ctatattaaa 4200
 aattaaaatt ttaatggcca ggcacggtgg ctcatgtccg taatcccgagc accttgggag 4260
 5 gccgaggcgg gtggatcacc tgaggtcagg agtttgaaac cagctctggcc accatgatga 4320
 aacctgtct ctactaaaga tacaaaaatt agccaggcgt ggtggcacat acctgtagtc 4380
 ccaactactc aggaggctga gacaggagaa ttgcttgaac ctgggaggcg gaggttgagc 4440
 tgagccgaga tcatgccact gcactgcagc ctgggtgaca gagcaagact ccatctcaaa 4500
 acaaaaaaaa aaaaaaagag attaaaaatt taatttttat gtaccgtata aatatatact 4560
 ctactatatt agaagttaaa aattaaaaaca attataaaag gtaattaaacc acctaatcta 4620
 10 aaataagaac aatgtatgtg gggtttctag cttctgaaga agtaaaagt ttggccacga 4680
 tggcagaaat gtgaggaggg aacagtggaa gttactgttg tttagacgctc atactctctg 4740
 taagtgactt aattttaacc aaagacaggc tgggagaagt taaagaggca ttctataagc 4800
 cctaaaaaaa ctgctaataa tgggtgaaagg taatctctat taattaccaa taattacaga 4860
 tatctctaaa atcgagctgc agaattggca cgtctgatca caccgtcttc tcattcacgg 4920
 15 tgcctttttt cttgtgtgct tggagatttt cgattgtgtg ttctgttttg gttaaaacta 4980
 atctgtatga atcctgaaac gaaaaatggt ggtgatttcc tccagaagaa tttagagtacc 5040
 tggcaggaag caggtggctc tgtggacctg agccacttca atcttcaagg gtctctggcc 5100
 aagaccagag tgcaaggcag aggcctgatg acccgaggac aggaaagctc ggatgggaa 5160
 gggcgatgag aagcctgcct cgttgggtgag cagcgcatga agtgccctta tttagccttt 5220
 20 gcaaaagattg ctctggatag catctggaag agcgccagc cgggaatgca aggagtcaga 5280
 agcctcctgc tcaaaaccag gccagcagct atggcgccca cccggcgctg tgccagaggg 5340
 agaggagtca aggcacctcg aagtatggct taaatctttt ttccacctga agcagtgacc 5400
 aaggtgtatt ctgagggaag cttgagttag gtgccttctt taaaacagaa agtcattgga 5460
 gcacctctc caagggaaaa ccagacgccc gctctgcggt catTTacctc ttctctctct 5520
 25 cctctctctg cctctcgctg ttctgatcgg gacagagtga ccccgctgga gcttctccga 5580
 gcccgtgctg aggacctct tgcaaaaggc tccacagacc cccgcctctg agagaggagt 5640
 ctgagcctgg cttaataaca aactgggatg tggctggggg cggacagcga cggcgggagt 5700
 caaagactta attccatgag taaattcaac ctttccacat ccgaatggat ttggatttta 5760
 30 tcttaatat ttcttaaat tcatcaata acattcagga ctgcagaaat ccaaaaggct 5820
 aaaacaggaa ctgagctatg ttgccaagg tccaaggact taataacat gttcagaggg 5880
 attttccgcc ctaagtactt ttatttggtt ttcataagg gtcttagggt gcaaggga 5940
 gtacacgagg agaggcctg gcggcaggc tatgacacg gcagggccac cggggagaga 6000
 gtccccggc tgggaggctg acagcaggac cactgacctt cctccctggg agctgccaca 6060
 35 ttgggcaacg cgaagcgccg cagctgcgt gtgactcagg accccatacc ggcttctctg 6120
 gccccccac actaaccag gaagtccagg agctctgaac cctgggaaac gaacatgacc 6180
 ctgacctgct tgcttccctg ggtgggtcaa ggttggtgca ggttggtgca ggaatggcc 6240
 atgtaaatta cagactctg ctgatgggga cgttctctc catcattatt catcttacc 6300
 ccaaggact gaattattc agcaacttct tgggtgtga caagccatga caaaactcag 6360
 40 tcaaaacacc actcttttac tagggccaca gagcacgsc cacacctg atatatagg 6420
 agtccaggag agatgaggct gctttcagcc accaggctgg ggtgacaaca ggggtgaac 6480
 agctgttcc tctagactag tagacctg caggcactcc cccagattct agggcctggt 6540
 tgcgtcttcc cgaggcgcc atctgacctg gagactcagc ctggggtgac aactgaggc 6600
 cagccctgtc tccacacct ccgctccag gcctcagct ctccagcagc ttcttaaac 6660
 45 ctgggtggg cgtgttccag cgtactgtc tcacctgtc cactgtgtct tgtctcagc 6720
 acgtagctcg caggttctc cctcacatgg ggtgtctgtc tcttcccca acactcacat 6780
 gcgttgaagg gaggagattc tgcgctccc agactggctc ctctgagcct gaacctggct 6840
 cgtggccccc gatgcagggt cctggcgctc ggctgcacgc tgacctccat ttccaggcgc 6900
 tccccgtctc ctgtcatctg ccggggcctg ccggtgtgtt cttctgttct tgtgtcctt 6960
 50 tccacgtcca gctgcgtgtg tctctgccc ctagggtctc ggggttttta taggcatagg 7020
 acggggcggt ggtgggcca ggcgctctg ggaaatgcaa catttgggtg tgaagtagg 7080
 agtgctgtc ctacactagg tccacgggca caggcctggg gatggagccc ccgccaggga 7140
 ccgccccctc tctgcccagc acttctctgc cccccctc ctggaacaca gaggggcagt 7200
 ttccacaagc actaagcatt ctcttcccaa aagacccagc attggcacc ctggacattt 7260
 55 gccccacagc cctgggaatt cagtgacta cgcacatcat gtacacact cgtccacga 7320
 ccgaccccc ctgttttatt ttaatagcta caaagcaggg aaatccctgc taaaatgtcc 7380
 gaggaaatg cgtttataa agcctgcagg catctcaagg gaattacgct gaggtaaac 7500
 tgccacctcc atgggatacg tacgcaacat gctcaaaaag aaagaatttc accccatggc 7560
 aggggagtggt ttaggggggt taaggacggt gggggcgcca gctgggggct actgcacga 7620
 60 ccttttacta aagcagttt cctgggtctg atggattgg ctcaattatg ggagactaac 7680
 cataggggag tgggggatgg ggaacccgga ggctgtgcca tctttgccat gcccagagt 7740
 cctgggcagg ataattgctc agagatgccc acgtcctgat tcccccaaac ctgtggacag 7800
 aacccgcccc gcccagggc ctttgcagg gtgatctccg tgaaggacct gaggctcggg 7860
 65 atccttcggg actacctgca ggcggaaaa gtaattcagg ggtctggga agaggcgggc 7920
 aggagggtca gaggggggca gcctcaggac gatggaggca gtcagctcga ggctgaaaag 7980

6 / 18

5 ggagggaggg cctcgagccc aggcctgcaa gcgcctccag aagctggaaa aagcggggaa 8040
 gggaccctcc acggagcctg cagcaggaaag gcacgggctg cccttagccc accagggccc 8100
 atcgtggacc tcggcctccc gtgccatagg agggcactcg cgctgccctt ctagcatgaa 8160
 gtgtgtgggg atttcagaaa gcaacaggaa acccatgcac tgtgaatcta ggattatttc 8220
 10 aaaacaaagg ttacagaaa catccaagga cagggctgaa gtgcctccgg gcaagggcag 8280
 ggcaggcacg agtgatttta tttagctatt ttattttatt tacttacttt ctgagacaga 8340
 gttatgctct tggtgccacg gctggagtg cgcggcatga tcttggtca ctgcaacctc 8400
 cgctccctgg gttcaagcaa ttctcgtgcc tcagcctccc aagtagctgg gatttcaggc 8460
 gtgcaccacc acaccggcct aattttgtat tttagtaga gatgggcttt caccatgttg 8520
 gtcaagctga tctcaaaatc ctgacctcag gtgatccgcc caccacagcc tcccaaatg 8580
 ctgggattac aggcattgagc cactgcacct ggcctattta accatttta aacttccctg 8640
 ggctcaagtc acaccactg gtaaggagtt catggagttc aatttccctt ttactcagga 8700
 gttaccctcc ttgtatattt tctgtaattc ttctgtagt ggggatacac cgtctcttga 8760
 15 catattcaca gtttctgtga ccacctgtta tcccatggga cccactgcag ggcagctgtg 8820
 gaggtctgag gcttcaggtc ccagtggggt tgccatctgc cagtagaaac ctgatgtaga 8880
 atcagggcgc aagtgtggac actgtcctga atctcaatgt ctcagtgtgt gctgaaacat 8940
 gtagaaatta aagtcctacc ctctactct actgggattg agcccccttc ctatcccccc 9000
 ccagggcgag aggagttcct ctcaactcctg tggaggaagg aatgatactt tgttattttt 9060
 cactgctggt actgaatcca ctgtttcatt tgggtggttg ttgttttgt tttagaggcg 9120
 20 ggtttcactc ttgttctca ggctggaggg agtgcaatgg cgcgatcttg gcttactgca 9180
 gcctctgctt cccaggttca agtgattctc ctgcttccgc ctccatttg gctgggatta 9240
 caggcacccg ccaccatgcc cagctaattt ttgtatttt tagtagagac ggggggtggg 9300
 ggggttcacc atgttgcca ggctggtctc gaacttctga cctcagatga tccacctgcc 9360
 tctgcctcct aaagtgtctg gattacaggt gtgagccacc atgcccagct cagaatttac 9420
 25 tctgtttaga aacatctggg tctgaggtag gaagctcacc ccactcaagt gttgtgggtg 9480
 tttaagccaa tgatagaatt tttttattgt tggtagaaca ctcttgatgt ttactactgt 9540
 gatgactaag acatcatcag cttttcaaa acacactaac tgcaccata atactggggt 9600
 gtcttctggg tatcagcaat ctctattgaa tgccgggagg cgtttcctcg ccatgcacat 9660
 30 ggtgttaatt actccagcat aatcttctgc ttccatttct tctcttccct cttttaaaa 9720
 tgtgttttct atgttggtct ctctgcagag aaccagtgtg agctacaact taacttttgt 9780
 tggaaacaa ttctcaaac ccctcttctg cctagtggca gagacaattc acaaacacag 9840
 ccctttaaaa aggtcttagg atcactaagg ggatttctag aagagcgacc tgtaattcta 9900
 agtatttaca agacgaggct aacctccagc gagcgtgaca gcccaggag ggtgcgaggc 9960
 35 ctgttcaaat gctagctcca taaataaagc aatttcctcc ggcagtttct gaaagttaga 10020
 aagggtacat ttaagggtgc gtttgttagc atttcagtg ttgcccagct cagctacagc 10080
 atccctgcaa ggctcggga gacccagaag tttctgccc ccttagatcc aaacttgagc 10140
 aacccggagt ctggattcct ggggaagtcct cagctgtcct gcggttgtgc cggggcccca 10200
 ggtctggagg ggaaccagtgg ccgtgtggct tctactgtg ggttggaagt cgggctcct 10260
 40 agctctcag tccgaggctt ggagccaggt gcctggacce cgaggctgcc ctccacctg 10320
 tgccggcgagg atgtgaccag atgttggtct catctgccag acagagtgc ggggccagg 10380
 gtcaaggccg ttgtggctgg tgtgaggcgc ccggtgcgcg gccagcagga gcgcctggct 10440
 ccatttccca ccttttctcg acgggaccgc cccggtgggt gattaacaga ttgggggtg 10500
 tttgctcatg gtggggacce ctgcgcgcct gagaacctgc aaagagaaat gacgggctg 10560
 45 tgtcaaggag cccaagtgc ggggaagtgt tgcagggagg cactccggga ggtcccgcgt 10620
 gcccgteccag ggagcaatgc gtctcgggt tctgtccccc cgcgctctac gcgcctcctg 10680
 cctcccttc acgtccggca ttctgtgtgc ccggagcccc acgccccgcg tccggacctg 10740
 gaggcagccc tgggtctccg gatcaggcca gcggccaaag ggtcgcgcga cgcacctgtt 10800
 cccagggcct ccacatcatg gccctccct cgggttacc caccagctag gccgattcga 10860
 50 cctctctccg ctggggccct cgtggcgct cctgcacct gggagcgcca gcggcgcgcg 10920
 ggcggggaa ggcggccag acccccgggt ccgcccggag cagctgcgct gtcggggcca 10980
 ggcgggctc ccagtggatt cgcgggcaca gacgcccagg accgcgctcc ccacgtggcg 11040
 gagggactgg ggaacccggg acccgtcctg ccccttcacc ttccagctcc gcctcctccg 11100
 cgcgagcccc gccccgtccc gacccctccc ggggtcccg cccagcccc tccgggcccc 11160
 55 cccagcccc ccccttctct tccgggccc cgcctctccc tccggcgcg agtttcaggc 11220
 agcgtgctgt cctgctgcgc acgtgggaag ccttgcccc ggcaccccc gcgatg 11276

<210> 4
 <211> 104
 <212> DNA
 60 <213> Homo sapiens

<400> 4
 gtgggctccc ccggggctcg cgtccggctg ggggtgaggg cgccggggg gaaccagcga 60
 catgcggaga gcagcgagg cgactcagg cgcttcccc gcag 104

7 / 18

<210> 5
<211> 8616
<212> DNA
<213> Homo sapiens

5

<400> 5

gtgaggaggt ggtggccgtc gagggccag gcccagagc tgaatgcagt aggggctcag 60
aaaagggggc aggcagagcc ctggctctcc tgtctccatc gtcacgtggg cacacgtggc 120
ttttcgctca ggacgtcgag tggacacggt gatctctgcc tctgctctcc ctccgtgcca 180
10 gtttcgataa acctacgagg ttacacctca cgttttgatg gacacgcggt ttccaggcgc 240
cgaggccaga gcagtgaaca gaggaggtg ggcgcggcag tggagccggg ttgcccggca 300
tggggaagaag tgtctggaag cacagacgct ctggcgaggg tgcctgcagg ttacctataa 360
tcctctctgc aatttcaagg gtgggaatga gaggtgggga cgagaacccc ctcttctctg 420
gggtgggagg taaggggttt gcagggtcac gtggtcagcc aatagtcagg tttgtgttta 480
15 agatttaatt gtgtgttgac ggcagggtgc ggtggctcac gccggtaatc ccagcacttt 540
gggaagctga ggcagggtga tcacctgagg tcaggagtgt gagaccagcc tgaccaacat 600
gggtgaaacc tatctgtact aaaaatacaa aaattagctg ggcaggtgtg tgtgtgcctg 660
taatcccagc tacttgggag gctgaggcag gagaatcact tgaacccagg aggcggaggc 720
tgcagtgagc tgagattgtg ccattgtact ccagcctggg cgacaagagt gaaactcctg 780
20 ctttaaaaaa aaaaagtgtt cgttgattgt gccaggacag ggtagaggga gggagataag 840
actgtctccc agcacagatc ctggtcccat cttaggtat gaagagggcc acatgggagc 900
agaggacagc agatggctcc acctgtctag gaaggacag tgtttgtggg tgttcagggg 960
atggtgctgc tgggcccctgc cgtgtcccca ccctgttttt ctggatttga tgttgaggaa 1020
cctccgctcc agcccccctt tggctcccag tgctcccagg cctaccctg gcagctagaa 1080
25 gaagtcccca tttcaccccc tccccacaaa ctcccagac atgtaagact tccggccatg 1140
cagacaagga ggggtgacct cttggggctc tttttttctt tttttcttct ttatggtggc 1200
aaaagtcata taacatgaga ttggcactcc taacaccgtt tctgtgttac agtgagaat 1260
tgctaactcg gcggtgttta cagcaggttg cttgaaatgc tgcgtcttgc gtgactggaa 1320
gtccctaccc atcgaacggc agctgctca cacctgctgc ggtcagggtg gaccacggc 1380
30 agtcagataa gcgtcatgca acccagtttt gctttttgtg ctccagcttc ctctgttag 1440
gagagtttga gttctctgat caggactctg cctgtcattg ctgttctctg acttcagatg 1500
agggtcacaat ctgccccctg cttatgcagg gagttagggc tgggtccccg gtgtccctgt 1560
cacgtgcagg gtgagttagg cgttgcccc aggtgtccct gtcacgtgta gggtagtga 1620
ggcgcgcccc ccgggtgtcc ctgtcccgtg cagcgtgatt gaggtgtggc ccccggtgt 1680
35 ccctgtcacg tgtagggtga gtgaggcgc atccccgggt gtccctgtca cgtgtagggt 1740
gagttagggc tgggtccccg gtgtccctgt cccgtgcagg gtgagttagg cactgtcccc 1800
gggtgtccct gtcacgtgca gggtagtga ggcgcggtcc ccgggtgtcc ctctcagggt 1860
tagggtagt gaggcgcggc cccagggtgt cctgtcacg tgtagggtga gtgaggcacc 1920
gtccctgggt gtccctccca ggtatagggt gagttagggc ctgtccccg gtgtccctgt 1980
40 cacgtgcagg gtgagttagg cgcggccccc ggtgtccct ctccagtgca gggtagtga 2040
ggcgctgtcc ctgggtgtcc ctgtctctgt tagggtagt gaggtctgt cccaggtgt 2100
ccttggcgtt tgctcacttg agcttctcc tgaatgtttg ctcttctat agccacagct 2160
gcgcgggttg cccattgcct gggtagatgg tgcaggcgca gtgctgggtc ccaagcctat 2220
cttttctgat gctcggctct cttgggtcac ctctccgttc cattttgtca cggggacacg 2280
45 ggactgcagg ctctcgcctc ccgctgcca ggcactgcag ccacagcttc aggtccgctt 2340
gcctctgttg ggcctggtt gctcaccagc tgcgccccac atgcatgctg ccaatactcc 2400
tctcccagct tgtctcatgc cgaggctgga ctctgggctg cctgtgtctg ctgccagctg 2460
ttgctggaga catcccagaa agggttctct gtgccctgaa ggaagcaag tccccacg 2520
ccctcactt gtccctgttt ctcccagct gccctctgct ttggccccc tgggtgggtg 2580
50 gcaacgcttg tcaccttatt ctgggcacct gccgctcatt gcttaggctg ggtctgcct 2640
ccagtgcgcc cctcacatgg attgacgtcc agccacagg tggagtgtct ctgtctgtct 2700
cctgtctcga gaccacgtg gagggccggt gtctccgcca gccttcgtca gacttccctc 2760
ttgggtctta gttttgaatt tcaactgatt acctctgacg tttctatctc tccattgtat 2820
55 gctttttctt ggtttattct ttcattcctt ttctagcttc ttagtttagt catgccttc 2880
cctctaagt ctgccttacc tgcacctgt gttttgatgt gaagtaatct caacatcagc 2940
cactttcaag tgttcttaaa atacttcaa gtgttaatac ttcttttaag tattcttatt 3000
ctgtgatttt tttctttgtg cagcgtgtgt tttgacgtga aatcattttg atatcagtga 3060
cttttaagta ttcttttagct tattctgtga ttcttttag cagttagtta ttgaaacact 3120
60 gtttatgttc aagatatgta gagtatcaag atacgttagg tattttaagt tatcatttta 3180
ttattgatt ctaactcagt tgtgtagtgg tctgtataat accaattatt tgaagtttgc 3240
ggagccttgc tttgtgatct agtgtgtgca tgggttccag aactgtccat tgtaaaattg 3300
acatccctgc aatagtggtg atgcatgttc actatataca gcttattaag gtccagtgca 3360
aagcttctgt ctccctctag atgcatgaaa ttccaagaag gaggccatag tccctcactt 3420
65 gggggatggg tctgttcatt tcttctcgtt tggtagcatt tatgtgaggc attgttaggt 3480
gcattgcagc ggtagaattt ttattcttct gatgagtga tcttctggag acttctatgt 3540

ctctagtaaat ctagtaaatc ttttttctaaa ttgctcttag tactgccaca ctgggcttct 3600
tttgattagt atttttctgc tgtgtctgtt ttctgccttt aatttatata tatatatata 3660
tttttttttt ttttgagaca gagtcttggc ctgtcgccca gggtagtgac agtgggtgta 3720
tcacaggcca gtgtaacttt taccttctgg cctgagccgt cctctcacct cagccctctg 3780
5 agtagctgga actgcagaca cgcaccgcta cacttggtta atttttaaat tttttctgga 3840
gacagggctt tgcgtgtgtt cccaggctgg tctcaaaact ttggactcaa gggatccatc 3900
tacctcggct tcccaaatg ctgaattaca ggcagtagcc accatgtctg gcctaatttt 3960
caacactttt atattcttat agtgtgggta tgcctgttta acagcatgta ggtgaatttc 4020
10 caatccagtc tgacagtcgt tgtttaactg gataacctga tttattttca tttttttgtc 4080
actagagacc cgcctgggtg actctgattc tccacttgcc tgttgcatgt cctcgttccc 4140
ttgtttctca ccacctcttg ggttgccatg tgcgtttcct gccagtggtg tgttgatcct 4200
ctcgttgccct cctggctact gggcatttgc tttattttct ctttgcttag tgttaccctc 4260
tgatcttttt attgtcgttg tttgcttttg tttattgaga cagtctcact ctgtcaccca 4320
ggctggagtg taatggcaca atctcggctc actgcaacct ctgcctcctc ggttcaagca 4380
15 gttctcattc ctcaacctca tgagtgcgtg ggattacagg cgcccaccac cacgctggc 4440
taatttttgt atttttagta gagataggct ttaccatgt tggccaggct ggtctcaaac 4500
tcctgacctc aagtgatctg cccgccttgg cctcccacag tgctgggatt acaggtgcaa 4560
gccaccgtgc ccgcatcacc ttgatctttt aaaatgaagt ctgaaacatt gctacccttg 4620
tcctgagcaa taagaccctt agtgtatttt agctctggcc acccccagc ctgtgtgctg 4680
20 ttttccctgc tgacttagtt ctatctcagg catcttgaca ccccacaag ctaagcatta 4740
ttaatattgt tttccgtgtt gagtgtttct gtgctttgc ccccgccctg ctttccctcc 4800
tttgttcccc gtctgtcttc tgtctcaggc ccgcccctg gggctcccct ccttgccttc 4860
tgcgtgggtc ttctgtcttg ttattgctgg taaacccag ctttacctgt gctggcctcc 4920
atggcatcta gcgacgtccg gggacctctg cttatgatgc acagatgaag atgtggagac 4980
25 tcacgaggag ggcggtcacc ttggcccgtg agtgtctgga gcaccacgtg gccagcgttc 5040
cttagccagt gagtgcagc aacgtccgct cggcctgggt tcagcctgga aaacccagg 5100
catgtcgggg tctggtggct ccgcggtgct gagtgtgaaa tcgcgcaaac ctgcggtgtg 5160
gcgccagctc tgacggtgct gcctggcggt ggagtgtctg ctctccctct tctgcttggg 5220
aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgtgagcc 5280
30 atgtggataa ttttaaaatt tctaggctgg gcgcggtggc tcacgcctgt aatcccagca 5340
ctttgggagg ccaaggcggt tggatcacga ggtcaggagg tcgagaccat cctggccaac 5400
atgatgaaac cccatctgta ctaaaaaac aaaaattagc tgggctggtt ggcgggtgcc 5460
tgtaatccca gctactcggg aggtctgagg aggagaattg cttgaacctg ggagtggaa 5520
gttgagtgga gccgacattg caccactgca ctccagcctg gcaacacagc gagactctgt 5580
35 ctcaaaaaaa aaaaaaaa aaaaaaaa aattctagta gccacattaa aaaagtaaaa 5640
aagaaaagggt gaaattaatg taataataga ttttactgaa gccagcatg tccacacctc 5700
atcatttttag ggtgttatg gtgggagcat cactcacagg acatttgaca ttttttgagc 5760
tttgtctgcg ggatcccggt tgtaggtccc gtgcgtggcc atctcggcct ggacctgctg 5820
ggcttcccat ggccatggtc gttgtaccag atggtgcagg tccgggatga ggtcgcagg 5880
40 cccctcagtg gctggatgtg cagtgtccgg atggtgcacg tctgggatga ggtcgcagg 5940
ccctgctgtg agctggatgt gtggtgtctg gatggtgcag gtcagggtgt aggtctccag 6000
gccctcgggt agctggaggt atggagtccg gatgatgcag gtcgggggt aggtcgcagg 6060
gccctgctgt gagctggatg tgtggtgtct ggatgggtga ggtcagggtt gaggtctcca 6120
ggccctcgggt aagctggagg tatggagtcc ggatgatgca ggtccgggtt gaggtcgcca 6180
45 ggcctctgct tgagctggat gtgtggtgtc tggatggtgc aggtctgggt tgaggtcacc 6240
aggccctgct gtgagctggg tgtgcggtgt ctggatggtg cagggtctgga gtgaggtcgc 6300
cagacgggtc cagaccatgc ggtgagctgg atatgcggtg tccggatggt gcaggtctcg 6360
ggtgaggttg ccaggccctg ctgtgagttg gatgtgggtt gtcgggatgc tgcagggtccg 6420
gtgtgaggtc accaggccct gctgtgagct ggatgtgtgg tgtctggatg gtgcagggtc 6480
50 ggggtgaagg tcgccaggcc cctgcttctg agctggatgt gtggtgtctg gatggtgcag 6540
gtctggagtg aggtcgcagg gccctcgggt agctggatgt gcagtgtcca gatggtgcag 6600
gtccgggggt aggtcgcagg accctcgggt gagctggatg tgcgggtgtc ggatggtgca 6660
ggtctggagt gaggtcgcca ggccctcgggt gagctggatg tatggagtcc ggatggtgcc 6720
ggctccgggt gaggtcgcca gacctgctg tgagctggat gtgcgggtgtc tggatggtac 6780
55 aggtctggag tgaggtcgcc agacctgct gtgagctgga tatcggtgtc ccggatggtg 6840
caggtcaggg gtgaggtctc caggccctcg gtgagctgga ggtatggagt ccggatgatg 6900
caggtccggg gtgaggtcgc caggccctgc tgtgaactgg atgtgcggcg tctggatggt 6960
gcaggtcttg ggtgtggtcg ccaggccctc ggtgagctgg aggtatggag tccggatgat 7020
gcaggtccgg ggtgaggtcg ccaggccctg ctgtgagctg gatgtgcggc gtcctggatg 7080
60 tgcaggtctg ggtgtggtc gccaggccct cggtagctg gaggtatgga gtcgggatga 7140
tgcaggtccg ggtgaggtt gccaggccct gctgtgagct ggatgtgctg tatccggatg 7200
gtgcagtcgg ggtgaggtc gccaggccct gctgtgagct ggatgtgctg tatccggatg 7260
gtgcaggtct ggggtgaggt caccaggccc tgcggtgagc tgggtgtgct ggtccgggtt 7320
gctgcaggtc cggggtgagt tcgccaggcc ctccggtgagc tggatgtgct ggtcccggt 7380
65 gtcgggatg tgacaggtcca ggtgaggtc gctaggccct tgggtgggctg gatgtgctg 7440

5 gtcaggatgg tgcagggtctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgcggt 7500
 gtcctcatgg tgcagggtctg ggggtgaggtc gccaggccctt tggtagagctg gatgtgtggg 7560
 gtcaggatgg tgcagggtccg gcgtgaggtc gccaggccctt gctgtgagct ggatgtgcgg 7620
 tgtctggatg gtgcagggtcc ggggtgaggt agccaaggcc ttcggtgagc tggatgtggg 7680
 gtgtccggat ggtgcagggtc cggggtgagg tcgccaggcc ctgagggttag ctggatatgc 7740
 ggtgtccgga tgggtcagggt ccggggtgag gtcaccaggc cctgcgggta gctggatgtg 7800
 cggtgtctgg atggtgcagg tccggggtga ggtcgccagg cctgtctgtg agctggatgt 7860
 gctgtatccg gatgggtgca gtcgggggtg aggtcgccag gccctgcagt gagctggtatg 7920
 10 tctgtatcc ggatgggtgca ggtctggcgt gaggtcgcca ggcctgcgg ttagctggat 7980
 atcggtgtgc ggatgggtgca ggtccggggt gaggtcacca ggcctgcgg ttagctggat 8040
 gtgcgggtgc cggatgggtg aggtctgggg tgaggtcgcc aggcctgtct gtgagctgga 8100
 tgtgtctgtat ccggatgggtg cagggtccggg gtgaggtcgc caggccctgc ggtgagctgg 8160
 atgtgtctga tccggatgggt gcagggtctgg cgtgaggtcg ccaggccctg cggtagagctg 8220
 gatgtgcagt gtacggatgg tgcagggtccg ggtgaggtc gccaggccctt gcggtgggtg 8280
 15 gtatgtgtgt tgtctggatg gtgcagggtcc ggggtgagtt cgccaggccc tgcggtgagc 8340
 tggatgtgtg gtgtctggat gctgcagggtc cggggtgagt tcgccaggcc ctcggtgagc 8400
 tggatatgctg gtgtcccggt gtccgaatgg tgcagggtcca ggtgaggtc gccaggccctt 8460
 tgggtgggctg gatgtgcggt gtccggatgg tgcagggtctg ggtgaggtc gccaggccctt 8520
 20 tggtagagctg gatgtgcggt gtccggatgg tgcagggtccg ggtgaggtc accaggccctt 8580
 cggtgatctg gatgtggcat gtccttctcg ttttaag 8616

<210> 6
 <211> 2089
 <212> DNA
 25 <213> Homo sapiens

<400> 6
 gtactgtatc cccacgccag gcctctgctt ctggaagtc tggaaacacca gcccgccctc 60
 agcatgcgcc tgtctccact tgcctgtgct tccctggctg tgcagctctg ggtctgggagc 120
 30 caggggcccc gtcacaggcc tgggtccaagt ggattctgtg caaggctctg actgcttgga 180
 gctcacgttc tcttacttgt aaaatcagga gttgtgcca agtggctctt aggggtttgta 240
 aagcagaagg gatttaaat agatggaaac actaccacta gcctccttgc ctttcctctg 300
 gatgtgggtc tgattctctc tctctttttt tttctttttt tgagatggag tctcactctg 360
 ttgccaggcc tggagtgcag tggcataatc ttggctcact gcaacctcca cctcctgggt 420
 35 ttaagcgatt caccagccct agcctcctaa gtactgtgga ttacaggcac ctgccaccac 480
 gcctggctaa tttttgtact tttaggagag acggggtttc accatgttgg ccaggctgggt 540
 ctcgaaacta tgacctcagg tgatccaccc accttggcct cccaaagtgc tgggtttaca 600
 ggctaagcca ccgtgccag ccccgattc tcttttaatt catgctgttc tgatgtaac 660
 tccaatctat tggatttagg tcatgagagg ataaaatccc acccacttgg cgactcactg 720
 40 caggggagcac ctgtgcagg agcacctggg gataggagag ttccaccatg agctaacttc 780
 taggtggctg catttgaatg gctgtgagat tttgtctgca atgttcggct gatgagagt 840
 tgagattgtg acagattcaa gctggatttg catcagttag ggcgggagc gctgggtctg 900
 gagatgccag cctggctgag cccaggccat ggtattagct tctccgtgtc ccgcccaggc 960
 tgactgtgga gggctttagt cagaagatca gggcttcccc agctccctg cacactcgag 1020
 45 tccctggggg gccttgtgac acccctatg ccaaatcagg atgtctgag agggagctg 1080
 cagcagacct cgtcagaggt aacacagcct ctgggctggg gaccccgagc tgggtgctgg 1140
 gccatttctt tgcatctggg ggagggtcag ggccttccct gtgggaacaa gtttaatacac 1200
 aatgcacctt acttagactt tacacgtatt taatgggtgt cgacccaaca tggctcattt 1260
 accagattt tggaaagaat ttaattgggg tgaccggaag gagcagacag acgtgggtgt 1320
 50 ccccaagatg ctcttgtca ctactgggac tgttgtctg cctggggggc cttggaggcc 1380
 cctcctccct ggacagggtg ccgtgcctt tctactctgc tgggctcgcg gcctgcggtc 1440
 agggcaccag ctccggagca cccgcccgc cagtgtccac ggagtgcag gctgtcagcc 1500
 acagatgccc aggtccagggt gtggccgctc cagcccccgt gccccatgg gtggttttgg 1560
 55 gggaaaaggc caagggcaga ggtgtcagga gactgggtggg ctcatgagag ctgattctgc 1620
 tcttggctg agctgccctg agcagcctct cccgcccctt ccatctgaag ggtgtggct 1680
 ctttctacct gggggtcctg cctggggcca gccttgggt accccagtgg ctgtaccaga 1740
 gggacaggca tctgtgtgg aggggcatgg gttcacgtgg ccccgatgc agcctgggac 1800
 caggctccct ggtgtgtatg gtgggacagt caccctgggg gttgaccgcc ggactgggag 1860
 60 tccccagggt tgactatagg accaggtgtc caggtgccct gcaagtagag gggctctcag 1920
 aggcgtctgg ctggcatggg tggacgtggc cccgggcatg gccttcagcg tgtgctgccc 1980
 tgggtgccct gagccctcac tgagtcgggt ggggctgtg gcttccctg agcttccccc 2040
 tagtctgtg tctggctgag caagcctcct gaggggctct ctattgcag 2089

10 / 18

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens

5

<400> 7
gtggctgtgc ttgtgtttta cttctttttt aaacagaagt gcgtttgagc cccacatttg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
10 cggcgccaac ccatttgtgc gcacagttag gtggccgagg tgccgggtgcc tccagaaaag 180
cagcgtgggg gtgtaggggg agctcctggg gcaggggacag gctctgagga ccacaagaag 240
cagccggggc agggcctgga tgcagcacgg cccgaggtcc tggatccgtg tccctgctgtg 300
gtgcgcagcc tccgtgcgct tccgcttacg gggcccgagg accaggccac gactgccagg 360
agccacccgg gctctgagga tccctggacct tgcctccagg ctccctgcacc ccacccctgt 420
ggctgcccgt gctgcggtga ccccgctcgc tgaggagagt gtgggggtgag gtggacagag 480
15 gtgtggcatg aggatcccggt gtgcaacaca catgcggcca ggaacccgtt tcaaacaggg 540
tctgaggaag ctgggagggg ttctagggtc cgggtctggg ttggctgggga cactggggag 600
gggctgcttc tcccctgggt ccctatgggt ggggtgggac ttggccggat ccacttttct 660
gactgtctcc catgtgttcc ccgccag 687

20

<210> 8
<211> 494
<212> DNA
<213> Homo sapiens

25

<400> 8
gtgggtgccc gggacccccg tgagcagccc tgctggacct tgggagtggc tgccctgattg 60
gcacctcatg ttgggtggag gaggtaacct tgggtggggc gcaggagagt caggtgaccc 120
tgtcactgtt gaggacacac ctggcaccta ggggtgaggc cttcagcctt tccctgacga 180
catggggccc actgtgcacc ctgactgccc gggctcctat tcccaaggag ggtccactg 240
30 gatccagatt tccgtcagag aaggaaccgc aacggctcag ccaccaggcc cgggtgcctt 300
gcacccagct cctgagccag gggctcctg tccctgaggc cagagagggg acacagcccg 360
ccctgccctt ggggtctgga gtgggtgggg tcagagagag agtgggggac accgccaggc 420
caggccctga gggcagaggt gatgtctgag ttctctcggt gccactgtca gtctcctcgc 480
ctccactcac acag 494

35

<210> 9
<211> 865
<212> DNA
<213> Homo sapiens

40

<400> 9
gtaagggtca cgtgtgatag tctgttccag gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagt cgtgtctgtg atgcgtttct gtgggtggagg tacttccatg atttacacat 120
ctgtgatatg cgtgtgtgtg acgtgtgtgt cgtgggtgcat gtatctgtgg cgtgcataat 180
45 tgtgtgtgtg gtgtgtgtgg cacgtgtgtg tccatggtgt gtgtgcctgt ggtgtgcatg 240
tgtgtgtgtc tgtgacacgt gcatgttcat gctgtgtgtc gcatgtctgt gatgtgccta 300
tttgtgtgtg gtgtgtgcat gtgtccgtga catatgcgtg tctatggcat ggggtgtgtgt 360
ggcccttgg ccttactcct tccctcctcca ggcattggtc gcaccattgt cctcacgctc 420
tcgggtgctg gtttggggag ctccacattc agggctcctc cttctagcat gggtgcccc 480
50 gtccctgtcac agggctgggc cttggagact gtaagccagg tttagagga gagtagggat 540
gctgggtgta ccttcctgga cccctggcac cccagggacc ccagctctggc ctatgccggc 600
tccatgagat ataggaaggc tgattcaggc ctgcctcccc gggacacact cctcccagag 660
cggccggggg ccttggggct cggcaggggt gaaagggggc ctgggcttgg gttcccaccc 720
agtggctcatg agcacgctgg aggggtaagc cctcaaagtc gtgccaggcc ggggtgcaga 780
55 ggtgaagaag tatccctgga gcttcgggtc ggggagaggc acatgtggaa acccacaagg 840
acctctttct ctgactttct gagct 865

60

<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens

65

<400> 10
tgtgggattg gttttcatgt gtgggatagg tggggatctg tgggattggg ttttatgagt 60
ggggtaacac agagtccaag gcgagcttcc tccctgtagt gggctcgcag gtgctccaac 120

agctttattg agggagacat atcttccctt gaactatggt cgggtttata gtaagtcagg 180
 ggtgtggagg cctcccttgg gctccctgtt ctgtttcttc cactctgggg tctgtgtggg 240
 cctgctgtgg tgtgtggcgg gtgggcaggg ctccaggccc tctctgtgtt cattggcctg 300
 gatgtggccc tggctacgct ccgtcccttg aattccccg cgagttggag gctttctttc 360
 5 tttctttttt tcttcttttt tttttttttt tgataacaga gtctcgctct tttttgccc 420
 ggctggagtg gtttggcgtg atcttggttc actgcaacct gtgtctctcg agttcaagca 480
 attctcttgc ctacgctccc caagtagctg gaattatagg cgcccaccac catgctgact 540
 aatttttcta atttttagtag agacgaggtt tctccatgtt ggccaggctg gtctcgaaact 600
 10 cctgacctca ggtgaccttc ccacctcggt ctcccaaagt gctgggatga cagggtgtgaa 660
 ccgcccgcgc cggcccgagac tgccttctcg cagcttccgt gagatctgca gcgatacgtg 720
 cctgcagcct tgggtgtgac aaacctccgtt tctcttcttc aggtctcgct aggggtcttt 780
 ccatttcatg actctcttca cagaagagtt tcacgtgtgc tgatttcccg gctgttccct 840
 gcgtaattgg tgtctgctgt ttatcgatgg cctccttcca tttcctttag gctttgttta 900
 ttgttgtttt tccggctcct tgaaggaaaa gtctcgatta tggatgtttg aactttcttt 960
 15 tctaaacaag catctgaagt tgccgttttc cctctaaagc agggatcccc agggccctgg 1020
 ctgtggagtg gcaccggtct ggggctgtt aggaacccgg cgacagcgg gaggtcagg 1080
 ggggtgtggg gagccagcgt tcccgcctga gccccgcccc tctcagatca gcagtgcat 1140
 gcggtgtctca gaggcgcaca caccctactg agaactgtgc gtgagagggg tctagattct 1200
 20 gtgctcctta tgggaatcta atgcctgatg atctgaggtg gaaccgtttg ctcccaaaac 1260
 catccccctc cccactgctg tctgtggaa aaactcgtct ccacgaaacc agtccctgg 1320
 accacaatgg ttggggaccc tgtgtctaaag acctgcttca gcagcctctc gtcagtgttg 1380
 atatatggc ttttctgtgt tgagtccaga ataattacgg atttctgtga tgccttcccg 1440
 cgacctcaga cccatgggct atttgtgggc gtgtgtcctg ctctctgggtt gggaaggggtg 1500
 25 caggcccat gtaccttctt gttactgcct tccagggttg tctcagggtg tgaatcgtac 1560
 tcatgtgtgt tttagccccc ggccttgcg ccagctcctg ggggtctggg aacatgtctga 1620
 agcacagagt caccgtgcgc gcttttctg gctcacaag ctccaggcct cctgtgtccg 1680
 tgttagtgtg tgtcacgtgc ctgctcacat cctgtcttgg ggacgcaggg gcttagcagg 1740
 tcccgtagta aatgacaagc gtcctggggg agtctgcaga ataggaggtg ggggtgcggg 1800
 30 tctctctccc gctcttctag actcttcttc tgctgtgtgt gtggctgcac ctgcatccct 1860
 gcaatccctc cagcactggg ctggagaggg ccgggagctc gactgccact tgtgccactg 1920
 gactgtggat ggcagtcggg caccgggggtc tgatgtgtgg tgactgtgga tggcggttgg 1980
 tcacaggggg ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtact 2040
 gtgctggggc gtcgtggggg ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga 2100
 35 tgtgtgtgact gtggatggcg gtcgtggggg ctgatgtgtg gactgtggat ggcggtcgtg 2160
 gggctctgat tgggtgactgt ggtggcagc cgtgggggtc gatgtgtgtg gactgtggat 2220
 ggcggtcgtg ggggtctgat tgggtgactgt ggtggcagc cgtgggggtc gatgtgtgtg 2280
 gactgtggat ggcggtcgtg ggggtctgat tgtgtgtgact gtggatggcg gtcgtggggg 2340
 ctgatgtgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg ctgtggatgg 2400
 40 cggtcgtggg gtcctgatgtg gtgactgtgg atggcggtcg tggggtctga tgtgtgtgtg 2460
 ctgtggatgg tgatcggtca cagggggtcgt atgtgtgtgt actgtggatg gcggtcgtgg 2520
 ggtctgatgt gtcgtgactg tggatgggtga tccgtcacag ggggtctgat tgtgtgtgact 2580
 gtggatggcg gtcgtggggg ctgatgtgtg gtgactgtgg atggcggttg gtcctggggg 2640
 tctgatgtgt ggtgactgtg gatggcgatc ggtcacaggg gtcctgatgt tgggtgactgt 2700
 45 ggatggcggt cgtgggggtc gatgtgtgtg gactgtggat ggcggtcgtg ggggtcgtg 2760
 tgtgtgtgact gtggatggcg gtcgtggggg ctgatgtgtg gactgtggat ggcggtcgtg 2820
 ggggtcgtg tgggtgactgt ggtggcggt cgtgggggtc gatgtgtgtg gactgtggat 2880
 50 ggcggttggg cccgggggtc tgatgtgtgt tgactgtgga tggcggtcgt ggggtcgtg 2940
 gtgggtgact tggatggcag tccgtggggg tgatgtgtgt tgactgtgga tggcggtcgt 3000
 ggggtcgtg gtgtgtgtgac tgtggatggc ggtcgtggg tctgatgtgt ggtgactgtg 3060
 gatggcggtc gtgggggtcgt atgtgtgtgt actgtggatg gcggtcgtg ggtcgtgtg 3120
 ggtgactgtg gatggcggtc gtgggggtcgt atgtgtgtgt actgtggatg gtcgtcgtg 3180
 acagggggtc gatgtgtgtg gactgtggat ggcggtcgtg ggggtcgtg tgtgtgtgact 3240
 55 gtggatggcg gtcgtggggg ctgatgtgtg gactgtggat ggcggtcgtg ggggtcgtg 3300
 tgtgtgtgact gtggatggcg gtcgtggggg ctgatgtgtg gactgtggat ggcggtcgtg 3360
 gtcacagggg tctgatgtgt ggtgactgtg gatggcggtc gtgggggtcgt atgtgtgtg 3420
 actgtggatg ggcggtcgtg ggtcgtgtgt ggtgtgactg tggatggcg tctgtgggtc 3480
 60 tgatgtgtgt tgactgtgga tggcggtcgt ggggtcgtg gtgggtgact tggatgggtg 3540
 tccgtcacag ggggtcgtg tgtgtgtgtg gcagggtggg tcccagggtg gtcgtgtgact 3600
 actttgcgtc ctccggcccc cggccccctg tcccaaaaca gaagcttccc aggcgctctc 3660
 tgggtctcat cccgccatcg ggtctggcgg cagggtccaca cgtcctgac ggaagaaaca 3720
 agtccccagc tctggccggg gcaggccaca tttgtggctc atgccctctc ctctgcccgc 3780
 ag 3782

<210> 11
<211> 980
<212> DNA
<213> Homo sapiens

5

<400> 11
gtctggggcac tgccttgacg ggttggggcac ggactcccag cagtgggtcc tcccctgggc 60
aatcactggg ctcatgaccg gacagactgt tggccctggg gggcagtggg gggaaatgagc 120
tgtgatgggg gcatgatgag ctgtgtgcct tggcgaaatc tgagctgggg catgccaggc 180
10 tgcgacagct gctgcattca ggcacctgct cacgtttgac tgcgcgccct ctctccagtt 240
ccgcagtgcc tttgttcatt atttgctaaa tgccttctct gccagttttg atcttgaggc 300
caaaggaaaag gtgtccccct cctttaggag ggcaggccat gtttgagccg tgtcctgccc 360
agctggcccc tcagtgtctg gtctgaggcc aaaggaaaac gtgtccccct cttaggagga 420
cgggccgtgt ttgagccacg ccccgctgag cggggcctctc agtgctgggt ctgtccacgt 480
15 ggcctctggg ccttttgacg atgtgggtctg tccacgtggc cctgtggctc tttgcagatg 540
cctgttagca cttgctcggc tctaggggac agtcgtgtcc accgcatgag gctcagagac 600
ctctggggca atttctcttg ctcccagggt ggggttgagg gtggcctggg ctgtcgggac 660
ccagaccctg tgcggcgag ctgggcagca actcctggat cacatatgcc atccgggcca 720
cggtagggctg tgtgggtgtg agcccagctg gaccacaggg tggcccagag gagacgttct 780
20 gtgtcacaca ctctgcctaa gcccatgtgt gtctgcagag actcggcccg gccagcccac 840
gatggccctg cattccagcc cagccccgca ctctcatcaca aacactgacc ccaaaaggga 900
cggagggtct tggccacgtg gtccctgctg tctcagcacc caccggctca ctcccactgt 960
tctcccgctc gctttcgcag 980

25

<210> 12
<211> 2485
<212> DNA
<213> Homo sapiens

30

<400> 12
gtgagtcagg tggccagggt ccattgccct gcgggtgggt gggcgggctg gcagggtctc 60
tgtcaccttc tctcctgccc ctccccact gnccttctgc ccggggccac cagagtctcc 120
ttttctggcc ccgccccctt ccggctcctg ggctgcagcc tcccgaggcc ccggaacat 180
ggctcggctt gcggcagccg gagcggagca ggtgccacac gaggcctgga aatggcaagc 240
35 ggggtgtgga gttgtcctg cgtggaggac gaggggcggt ggggtgtgtc gggtcagggt 300
tgcgcccagc gtttgagcct gcagcttctc agctccaagt tactactgac gctggacacc 360
cggctctcac acgcttgtat ctctctctcc cgatacaaaa ggattttatc cgattctcat 420
tctgttccct gtcgtgtgac ccccgcgagg gcgcgggctc ttctctctgt gactagattt 480
cccatctgga aagtgccggg ttgaccgtgt agtttctcct tctcgggggg cctgtggtgg 540
40 ccatggggca ggcggccctg gagagctgcc gtcacacagc cactgggtga gccacactca 600
cgtgtgtaga gccacagtgc ctggtgccac atcacgtcct ctggatttta agtaaaacca 660
cacacctccc ggcaggcatc tgcctgcgac cctgtgtgtg cctggggaga gtggtagcac 720
ggaggaaatt cgtgcacact caaggctcat agcaaggcca tccgcagcca ggtggaaact 780
ggaggcctct ctctgggacg gtctccagcg gataaaggac tgtgcacagc ttcggaagct 840
45 tttatttaaa aatataacta ttaattattg cattataagt aatcactaat ggtatcagca 900
attataatat ttattaaagt ataattagaa atattaagta gtacacacgt tctggaaaaa 960
cacaaattgc acatggcagc agagtgaatt ttggccgagg gacacgtgtg cacatgtgtg 1020
taagcggccc ccaggccccc agaattcgtc gacaaagtca cctccccaga gaagccacca 1080
cgggcctcct tctgtgtcgt gaattttatt aagatggatc aagtcacgta ccgtccacgt 1140
50 gtggcagggc tttggggaat gtgaggtgat gactgcgtcc tcatgcccctg acagacagga 1200
ggtagactgt tctgtcctgt ccctaggaca cggacaggcc cgaagctcta gtccccatcg 1260
tggctcaggt tggcctctga ataaaaacgt ctcaaaaacc tgttggccca aaaactaaga 1320
acagagagag tttcccatcc catgtgctca caggggcgta tctgcttgcg ttgactcgct 1380
gggctggcgg gactcctaga gttgggtcgt gtgcttctgt gcaaaaagtg cagtcctctt 1440
55 gcccatcact gtgatattct caccagcaag gaaagcctct tttcttttct tctctttttt 1500
ttttttgaga cggaaacgtc ctgttctctg cctgggcttg agtgagtggt cgcgatctca 1560
actcactgca acctccgccc cccgggttcc agcatttctc ctgcctcagc ctcccagca 1620
gctgagatta caggcaccca cccctgcgc ctggctaatt tttgtatttt tagtagagag 1680
gggtttttgc catgttggtc aggtgtgtct cgaactcctg acctcagggt atccacccac 1740
60 ctccgctccc caaagtgtgt ggattacagg tgtgagccat cacgcccagc cggaaagcct 1800
ctttttaagg tgaccaccta tagcgttctc cgaaaataac aggtcttgtt tttgcagtag 1860
gtctgcaagc tctcttagca acaggagtggt cgtcctgtgt gctctgggga tggctgaggg 1920
tcgctgggca gccatgcctt ctgtgtgcac ctttaggttc cagggggcta tctgtctctc 1980
actgtttgtc tgaaaacgca ccttggcat ccttgttttg agagtctctg ctctcctgtt 2040
65 gtcattgctga aactaggggc aaggtgtgtat ccgttggcgc gcagcggcta catgtagggt 2100

catgagtcctt tcaccgtgga caaattcctt gaaaaaaaaa aaaggagtcc ggtaaagcat 2160
tcattccggg tcaagtgtct ggttctgtga ataaactcta agatttaaga aaccttaatg 2220
aaagaaaacc ttgatgattc agagcaagga tgtggtcaca cctgtggctg gatctgtttc 2280
agccgccccca gtgcatggtg agagtgggga gcagggtattg tttgttcaga ggtctcatct 2340
5 ggtatgtttc tgagggtgtt gccggctgaa tggtagacgt gtcgtttgtg tgatgaggt 2400
tctgtgtctg tgtgtggctc ggtttgagtg tacgcatgtc cagcacatgc cctgcccgtc 2460
tctcacctgt gtcttcccg cccag 2485

<210> 13
10 <211> 1984
<212> DNA
<213> Homo sapiens

<400> 13
15 gtgaggcctc ctcttcccca ggggggcttg ggtgggggtt gatttgcttt tgaatgcattc 60
agtgttaata ttcttggtgc tctggagacc atgactgtct tgccttgagg aaccagacaa 120
ggttgcagcc ccttcttggg atgaagccgc acgggagggg ttgcacagcc tgaggactgc 180
gggtctccacg caggctctgt ccagcggcca tgtccagagg cctcagggct cagcaggcgg 240
gaggggccgct gccctgcattg atgagcatgt gaattcaaca ccgaggaagc acaccagctt 300
20 ctgtcacgtc acccaggttc cgttaggggc ctgtgggaga tggggctggt gcagcctgag 360
gccccacatc tcccagcagg ccttcgacag gtggcctgga ctgggcgcct cttcagccca 420
ttgcccattc cacttgcatg ggttctacac ccaaggacgc acacacctaa atatcgtgcc 480
aacctaagt ggttcaatc agctggcttt tattgacagc agttactttt ttttttttaa 540
tactttaagt tctagggtag atgtgcacga cgtgcagggt agttacatat gtatacatgt 600
25 gccatgttgg tgtgtgtcac ccattaatc atcatttaca ttaggtatat ctctaatgc 660
tatccctccc cactcccccc atcccatgac agggcctggt gtgtgatgtt cccacccctg 720
tgtccaagtg ttctcattgt tcagttccca cctgtgagtg agaacatgtg gtgtttggtt 780
ttctttcctt gcaatagttt gctcagagtg atgggtttcca gcttcgtcca tgtccctaca 840
aaggacatga actcatcctt ttttatgact gcatagtatt ccgtgggtga tatgtgccac 900
30 attttcttaa tccagcttat catcgatgga catttgggtt ggttgcaagt ctttctact 960
gtgaatagtg ccgcaataaa catacgtgtg catgtgtctt tatagcagca tgatttataa 1020
tcctttgggt atatacccg taatgggatg gctgggtcaa atggtatttc tagttctaga 1080
tccttgagga atcaccacac tgtcttccac aatgggtgaa ctagtattaca ctcccacaa 1140
cagtgtaaaa gtgttctggt gctggagagg atgtggacag cagttatttt tttatgaaaa 1200
35 tagtatcact gaacaagcag acagtttagt aaggatgctt cagggaagcct gcaggccaca 1260
cagccatttc tctcgaagac tccgggtttt tctgtgcat cttttgaaac tctagctcca 1320
attatagcat gtacagtgga tcaagggtct tcttcattaa ggttcaagtt cttagtgaa 1380
ataagtttat gtaacagaaa caaaaatttc ttgtacacac aacttgctct gggatttgga 1440
ggaaagtgtc ctgcagctgg cggcacactg gtcagccctc tgggacagga tacctctggc 1500
40 ccattggtcat ggggcgctgg gcttgggctt gaggggtcaca cagtgcacca tgcccagctt 1560
cctgtggata ggatctgggt ctcggatcat gctgaggacc acagctgcca tgctggtaaa 1620
gggcaccacg tggctcagag ggggcgaggt tcccagcccc agctttctta ccgtcttcag 1680
ttattttcc ctaagagtgt gagaagtggg gccgcgctt atggccttcg ttcgtcttca 1740
gctggcacag aattgcacaa gctgatggta aacactgagt acttataatg aatgaggaat 1800
45 tgctgtagca gttaactgta gagagctcgt ctgttggaag gaaatttaag tttttcattt 1860
aacgccttgg gagaatgtta ctttatttat ggtgtgtgaa attgtttgac attcagtcct 1920
tcgtagacag atactacgta aaaagtgtaa agttaacctt gctgtgtatt ttcccttatt 1980
ttag 1984

<210> 14
50 <211> 1871
<212> DNA
<213> Homo sapiens

<400> 14
55 gtgaggcccg tgccgtgtgt ctgtggggac ctccacagcc tgtgggcttt gcagttgagc 60
cccccggtgc ctgcccctgg caccgcagcg ttgtctctgc caagtccctt ctctctgccg 120
gtgctggatc cgcaagagca gaggcgcttg gccgtgcacc caggcctggg ggcgcagggg 180
caccctcggg agggagtggg taccgtgcag gccctgttcc tgcaagagac caccaggtt 240
60 acacacgtgg tgaatgcagg cggtagcctg gctcctgctg ctctttggaa agtcaagagt 300
ggcggtcctt gggggcccg tgagacccc aggaagctgt cacaggccct gcaggggcga 360
ggcggcagcc tcttccccag ggtgcacctg agcctgcgga gagcaggagc tgcagagtga 420
gctgcccac agcgttcgct gcggtcacgt tccctgcgtg ggttgtttgg gatcggtggg 480
agaatttga tttgtgagt gctgctgtct tgaaccaggg agatggctag gagtgggttt 540
65 cagagttgat ttttgtgaat caaactaaaa tcaggcacag gggacctggc ctacgacag 600

14 / 18

5 gggattgtcc aatgtggtcc ccctcaaggg cgccccacag agccgggtggg cttgttttaa 660
 agtgcgattt gacgagggac gagaacacctt gaaagctgta aaggggaaccc tcagaaaatg 720
 tggccgcag ggggtggttc aggtgctttg ctgggtctgtg tttgtgaaaa cccattttga 780
 cccgccctcc aagtccaccc tccaggtcca ccctccaggg ccgccctggg ctgggggtat 840
 10 gcctggcggt ccttgtgccc cagcccggag cacagcaggc tgtgcacatt taaatccact 900
 aagattcact cggggggagc ccaggtccca agcaactgag ggctcaggag tcttgaggct 960
 gctgagggga cagagcagac ggggaacgct gcttctgtgt ggcaagtcc tgaggggtgt 1020
 ggccagggag gtggctcaga gtgtatgttg gggccccacc gggggcagaa ctctgtctct 1080
 gatgagtcgg cagccatgta acaggaaggg gtggccacag ggagctggga atgcaccagg 1140
 15 ggagctgcgc agctggccga ggtcccaagg ccaggccaca ggaagggcag ggggacgccc 1200
 gggggccacag cagagggcgc aggaaggaa ggggatgccc aggccagagc agaggctacc 1260
 gggcacaggg gggctcccctg agctgggtga gcgaggctca tgactcggcg agggaaacctc 1320
 cttgacgtga agctgacgac tgggtgttgc cagctcacag cccagccagg tcccgcgcct 1380
 gagcaggaac tcagaacctt cccctttgtc taaagcacag cagatgcctt cagggcatct 1440
 20 aggagaaaa acggcaaatg gttgagaaac gtcttaaaag aagggtggat ggtggcaatt 1500
 tcttgtccag attttagtct gcccggacc acagatgagt ctataacggg attgtggtgt 1560
 tgccatgggg acacatgaga tggaccatca cagaggccac tggggctgca cctcccatct 1620
 gagtctctgg tgtccgggt ccaggccagg tcttctcatg ctacacctacc tgtccctgcc 1680
 gggagacagg gaaagcaccc cgaagtctgg agcagggctg ggtccaggct cctcagagct 1740
 25 cctgccaggc ccagcacctt gctccaaatc accactcttc tgggttttcc caaagcattt 1800
 aacaagggtg tcaggttacc tcctgggtga cgggcccgca tctgggggtg gacattgccc 1860
 ctctgcctta g 1871
 <210> 15
 <211> 3801
 <212> DNA
 <213> Homo sapiens
 <400> 15
 30 gtgagcgac ctggccggaa gtggagcctg tgcccggctg gggcagggtg tgctgcaggg 60
 ccgttgcgtc cactctgct tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120
 ggccacaggg tgcccctcgt cccatctggg gctgagcaga aatgcactct tctgtgggag 180
 tgaggggtgt cacaacggga gcagttttct gtgtattttt ggtaaaagga aatggtgcac 240
 35 cagacctggg tgactgagg tgtcttcaga aagcagctctg gatccgaacc caagacgccc 300
 gggccctgct gggcgtgagt ctctcaaac cgaacacagg gggcctgctg ggcagtgagt 360
 cctctgaacc cgagaccctg gggccctgct gggcgtgagt ctctccgaac ccagagactt 420
 cagggccctt ttgggcgtga gtctctccgc tgtgagcccc acactccaag gctcatccac 480
 agtctacagg atgccatgag ttcagtatca cgtgtgaccc atcaggggac agggccatgg 540
 40 tctggggggg gtctctacaa aattctgggg tcttgtttcc ccagagcccg agagctcaag 600
 gccccgtctc aggtctcagac acaaatgaat tgaagatgga cacagatgca gaaatctgtg 660
 ctgtttcttt tatgaataaa aagtatcaac attccaggca gggcaagggt gctcacacct 720
 ataataccag cactttggga ggcgaggtg ggtggatcac ttgagggcag gagtttgagg 780
 45 caaacctaac caacatagtg aaattccatt tctacttaaa aaatacaaaa attagcctgg 840
 cctggtggca cagcctgta gtcccgccta tgcgggaggc tgaggcagga gaatcatttg 900
 aacccaggag gcagaggttg cagtgaagcc agatcacacc actgcactcc agcctgggca 960
 acagagttag acttcatctt aaaaaaaaaa aaaaaagtat cagcattcca aaaccatagt 1020
 ggacaggtgt ttttttatcc tgcctctcga taatatctac tgggtgctgtg ctagaggccg 1080
 50 gaactggggg tgcccttcctc tgaagggcac accttcatgg gaagagaaat aagtggtgaa 1140
 tgggtgttaa accagaggtt taaactgggg tcctgtcgtt ctgagttaac agtccagatc 1200
 tggactttgc ctctttccag aatgctccct ggggtttgct tcatggggga gcagcaggtg 1260
 tggacacctt cgtgatgggg gagcagcagg tgcagacgcc ctcatgatgg gggagtggca 1320
 ggtgcagaca cctctgtgca tgggtgccag catgtccctg ttgcagctcc ctccccacaa 1380
 55 ggatgccggg ctctctgtgt ccccacagtc cctgcttccc tctcacagcc ttacctggtc 1440
 ctggccctca ctggctttgt ctgcatgatt tccacatttc ctgggctccc agcacctctt 1500
 cgctctctcc aggcacctct gcagtgtcgg ccataccagt cagctgtgaa ctgtccactg 1560
 cttattttgc tccccatgaa atgtattttt taggacaggc acccttgggt ccagcctctg 1620
 gcacagcatc agtgaatggt attgaaggac aaaggacaga caaacaatc aggaaaaatg 1680
 gttctctcta aacacattgc aaagccacag aggtagtgc aggatgggtg ggcacaggt 1740
 60 catcagatgt gggccaatg ccagaatatt ctgtgtctcc aaaggccact tggtcagagt 1800
 gtgtgactgc agagggtggt ctaaaagctc agcagtggag gcagtgggtc gccatactca 1860
 ggggtgaact acatcctctg tgtctgaagt atacagcaga ggcttgaagg gcatctggga 1920
 gaagaaaaca ggcaaaatga ttaagaaaaa tgaagaaagg aaagtggtaa gatgggaatt 1980
 ttcttgtcca gattttatgc tcccaaacca cagctcagat ggtagaatgt ggtcagaact 2040
 gatggacaga acaatagAAC aaacgggaag ccctatctct cagaaacgtg tgttaatgtg 2100
 65 gtatgtggca cagctgatgg aaaagagagt gtgtgtgtaa tttttttttc tgagaaaaat 2160

15 / 18

5 gactggaagc aaataagttg tgtctttaca gcatatacca gagcagattc taggtagaag 2220
 aggagacaca tgcaaaacac accagcaaca gaaataaaac aaaagactca aaggggaaggg 2280
 aggtgaacgt tccctggttt ggtgttgggg aaggacacac agggaggcgg atgaaaccag 2340
 tgaggcaacg ggcattgctt tcaactgcaga gaaactcagc ttgcctgagc cacagtgaag 2400
 atggccattc cctggagcgt ttgtgcacgt gatttattta aggcgccttg tgaggtcctg 2460
 cacattcatc ctctcacttt gttctcctaa ccacctgaga ggtagaggag gaaaggctcc 2520
 aggggagcag ccgccccttg tcacccagct ggcaaaaggg atgcatgatt gcagcctggc 2580
 ctctgtctcc ggggcccctg ctctgccga ggacccca caagtcagac ccataggctc 2640
 10 aggggtgagcc ggagcccaag gtcgtgttgg ggatggctgt gaaagaagaa atggacgtct 2700
 gatgcacact tgggaaggtc ctaccagcag cgtcaaagaa atgcatgtga aactgacagc 2760
 gagaccatc cctcaaagaa acgcacgtga aactgatggc gagacctgtc cccatccctc 2820
 atgctggctc cttttctggg cttgccaa gaagcatca ggttgaggca agctggaaa 2880
 acttttcttg aaagcagctt gtttgcattg aagtccctac aatgtcctgt gtcttccag 2940
 taattccact tctgaagtga ccagacatta tcacgggtct tatttaccat ttccagtgtt 3000
 15 ccaggcaggg ggacttgcca cagcaagtca cgaacctgcc caaatacagg gctaaggaga 3060
 tattatgcat caaaaaactt gctctgcat taaacatttt tcaaagaatt tttgaagaat 3120
 gtttaattgg acaaaacgtt tatttcaatg tagcagtgtt caaagctgga tgtaaaagaa 3180
 cacacccag gagcctgccc tgaatgtcat gtgtgttcat ctttggacat ggacatacat 3240
 20 gggcagtgag tgggtgtgag gccctggagg acatcggttg gatgcctcca tctgccccct 3300
 ctggagacac catgtgtgcc acgtgcactc actggagccc tgtttagctg gtgccacctg 3360
 gctcttccat ccttgagatt caaacacagt gagattcccc acgcccact cagtgttctc 3420
 ccacaaaaaa cctgagtcac acctgtgttc actcgaggga cgcccgggag ccagggtctc 3480
 acagtttatt atgtgttttt ggctgagtta tgtgcagatc tcatcagggc agatgatgag 3540
 25 tgcacaaaaa cggccgtgag aggtttggat acactcaaca tcaactagcca ggtcctggtg 3600
 gagtttggtc atgcagagtc tggatggcat gtagcatttg gtagtccatg agtgagcacc 3660
 cagcccccct gggctgcagc gcatgccccca ggcaggacaa ggaagcgga ggaaggcagg 3720
 aggtctcttg gagcaagctt tgcaggaggg ggctgggtgt ggggcaggca cctgtgtctg 3780
 acattcccc ctgtgtctca g 3801

30 <210> 16
 <211> 880
 <212> DNA
 <213> Homo sapiens

35 <400> 16
 gtgagcaggc tgatgggtcag cacagagttc agagtccagg aggtgtgtgc gcaagtatgt 60
 gtgtgtgtgt gtgcgcgctg gccctgcaagg ctgatggtga ctggctgcac gtaagagtgc 120
 acatgtacgc atatacacgt gagcacatac atgtgtgcat gtgtgtacat gaaggcatgg 180
 40 cagtgtgtgc acaggtgtgc aagggcacaa gtgtgtgcac atgcgaatgc acacctgaca 240
 tgcattgtgt ttctgtgcaca gtcgtgtggg cattcacgtg aggtgcatgc gtgtgggtgt 300
 gcagtgtgag tagcatgtgt gcacataaca tgtattgagg ggtcctcgtg ttcccccg 360
 taggtcctca gcaccagtgc cactccttac aggatgagac ggggtcccag gccttgggtg 420
 gctgagggtc tgaagtgcga gccctgaggg cattgtccca tctgggcatc cgctccact 480
 45 cctctcctg tgggcttctg tgtccactcc cctctcctg tgggcattta catccactcc 540
 actccctctc tctgtgggc atccgcgtcc actcccccct tctgtgggca tctgcgtcca 600
 cctccctct ctgtgggcat ttgcgtccac tccctctcct gggtccttcc tgtcttggcc 660
 gagcctcggg ggcaggcaga tgacacagag tcttgactcg cccagggtgg ttgcagctgt 720
 ccgggtgagg gccaggcccg atttacttgg gaagagggat agtttcttgt caaaatgttc 780
 50 ctctttcttg ttccatctga atggatgata aagcaaaaag taaaaactta aaatcccaga 840
 gaggtttcta cgtttctca ctctttcttg gcgactctag 880

55 <210> 17
 <211> 3186
 <212> DNA
 <213> Homo sapiens

60 <400> 17
 gtgagccgcc accaaggggg gcaggccag cctccaggga cctccgcgc tctgtctacc 60
 tctgaccggg ggccttcacct tggaaactcct gggttttagg ggcaaggaa gtcttacgtt 120
 ttcatgtgt ctgctgcctg tgcacagttc tgttcgcgtg gctctgtgca aagcacctgt 180
 tctccatctc tgggtagtgg taggagccgg tgtggcccca ggtgtcccca ctgtgcctgt 240
 gcaactggccg tgggacgtca tggaggccat cccagggcag caggggcatg gggtaagag 300
 atgtttatgg ggagtcttag cagaggaggc tgggaagggt tctgaacagt agatgggaga 360
 65 tcagatgccc ggaggatttg gggctcagc aaagaggggc gaggtgggtg caggtgaggg 420
 tcgctggccc cacccccggg aagggtgcagc agagctgttg ctccccacac agcccggcca 480

16 / 18

gcacctgtgc tctgggcatg gctgtgctcc tggaaacgttc cctgtcctgg ctggtcaggg 540
 ggtgccccctg ccaagaatcg acaactttat cacagagggg agggccaatc tgtggaggcc 600
 acagggccag cttctgctcg gagtacgggc aggtgggtggc acaagcctcg gggctgtacc 660
 aaagggcagc cgggcaccac agggccgggc ctccacctca acaggcctcc cgagccactg 720
 5 gtaggtgaat gccaggaggc cgaagccctc gccccatgag ggctgagaag gagtgtgagc 780
 atttgtgtta ccagggcccg aggtctgcgcg aattaccgtg cacacttgat gtgaaatgag 840
 gtcgtcgtct atcgtggaaa ccagcaagg gctcacggga gaggttttcca ttacaaggtc 900
 gtaccatgaa aatgggtttt aacccgagtg cttgcgcctt catgctctgg caggaggaggc 960
 10 agagccacag ctgcatgtta ccgcttttgc accagctcca gaggcttggg accaggctgt 1020
 ctcagttcca ggggtgcgtcc ggctcagacc gccctcctct ctgcctcttc tctctgcctc 1080
 aaatcttccc tcgtttgcat ctccctgacg cgtgcctggg ccctcgtgca agctgcttga 1140
 ctcccttccc gaaacccctg ggggtgtgtg gatacaggtg ccactgagga ctggagggtg 1200
 ctgacactgt ggttgacccc agggctccagc tggcgtgctt ggggcctcct tggggccatga 1260
 tgaggtcaga ggagttttcc cagggtaaaa ctccctggga actcccaggg ccattgtgacc 1320
 15 tgccacctgc tcctcccata ttcatctcag tcttgcctc atttcccac cagggtctct 1380
 agctccgagg agctccgta gagggcctgg gctcagggca gggcggtga gtttcccac 1440
 ccattgtggg acccttgggt agtcgcttga ttgggtagcc ctgaggaggc cgagatgcga 1500
 tggggccagg gccgtttcca aacacagagt caggcacgtg gaaggccca gaatcccctt 1560
 20 ccttcgaggc aggagtgagg gaacggagag ctggggcccc atttcacggc agccaggctg 1620
 cagtgggcga ggtgtgtgtg gtccacgttg cgctgggggc ggggtctgat tcaaatccgc 1680
 tggggctcgg ccttccctgg ccgtgtgtgg cgcgctcca caggggcttg ggggtggacg 1740
 cccgacctct agcaggtggc tatttctccc ttgggaagag agccctcac ccattgtgag 1800
 tgtttccctc ctgggtcagg agcgtggccg tgtggcaacc ccgggacctt aggtctattt 1860
 atttgtttaa aaacattctg gccctggctt ccgttgttg taaatgggga aaagacatcc 1920
 25 caccctcaga gagtactga gaggtgaaa ccgggtgctt ggcttgactg gtgtgactc 1980
 aggtcattcc agaagtggct cagggaagtca gtgagaccag gtacatgggg ggctcaggca 2040
 gtgggtgaga tgaggtacac ggggggctca ggcagtggtt gaggccaggt acatgggggg 2100
 ctccaggcact ggtgtgagat aggtacacgg ggggctcagg cagagggtca gaccagggtac 2160
 acgggggctc tgatcacacg cacatatgag cacatgtgca catgtgctgt ttcatggtag 2220
 30 ccaggctctg gcacacctgc cccaaagtcc cagggaagct agaggccaaa gatggaggct 2280
 gacagggctg gcgcgggtgg tcacacctgt agtcccagca ctttgggagg ccgaggcgag 2340
 aggatccctt gagcccagga gtttaagacc agcctgagca acatagtaga accccatctc 2400
 tatgaaaaat aaaaacaaaa attagctgaa catgggtggt tgcgcctgta gttccaatac 2460
 35 ttgggagggt gaagtgggag gatcactga gcccaggagg tggaaagctgc agtgagctga 2520
 gattgcacca ctgtactgca gcctgggtga cagagtga gaacctctca acaacaacaa 2580
 agaagactga caaatgcagt ttcttgga aaacatttta gtaggaaact aaacctacaca 2640
 cagaagccaa gtcgggtgtc cggtgtcagt gagatgagat gatgggtcct cacaccatca 2700
 cccagagccc aggggtttatg caccacaggg gcgggtgggt cagaagggat gcgcaggacg 2760
 40 ttgatatacg atgacatcaa ggttgcctga cgaagggcag gattcatgat aagtacctgc 2820
 tggtagacaa ggaacaatgg ataaactgga aaccttagag gccttcccgg aacagggggt 2880
 aatcagaagc cagcatgggg ggctggcacc caggatggag ctgcttcagc ctccacatgc 2940
 gtgttcatac agatgggtgca cagaacgca gtgtacctgt gcacacacag acacgcagct 3000
 actcgcacac acaagcacac acacagacat gcattgcacgc atccgtgtgt gtgcacctgt 3060
 gcccattagg aaacctatgc atgtgcattc atgcacgcac acaggcaccg gtggggcccat 3120
 45 gcccacaccc acgagcaccg tctgattagg aggcctttcc tctgacgctg tccgccatcc 3180
 tctcag 3186

<210> 18
 <211> 781
 50 <212> DNA
 <213> Homo sapiens

<400> 18
 55 gtatgtgcag gtgcctggcc tcagtggcag cagtgcctgc ctgctgggtg tagtgtgtca 60
 ggagactgag tgaatctggg cttagggaagt ccttaccctt tttcgcatca ggaagtgggt 120
 taacccaacc actgtcaggc tcgtctgccc gccctctcgt ggggtgagca gagcacctga 180
 tggaaaggac aggagctgtc tgggagctgc catccttccc acctgtctt gcctggggaa 240
 gcgctggggg gcctgggtct tcctgtttgc cccatgggtg gatttggggg gcctggcctc 300
 60 tcctgtttgc cctgtgggtg gatttgggtg tctcccgctc atggcactta gggcccttgt 360
 gcaaaccacg gccaaagggt taggaggagg ccaggccca gctacccac ccctctcagg 420
 agcagaggcc gcgtatcacc acgacagagc cccgcgccc cctctgcttc ccagtcaccg 480
 tcctctgccc ctggacactt tctccagcat caggagggtt tctgatccgt ctgaaattca 540
 agccatgtcg aaacctcggt cctgagctta acagcttcta ctttctgttc tttctgtgtt 600

17 / 18

gtggaaatctt cacctggaga agccgaagaa aacatttctg tctgtactcc tgcggtgctt 660
gggtcgggac agccagagat ggagccaccc cgcagaccgt cgggtgtggg cagctttccc 720
gtgtctctctg ggaggggagc tgggctgggc ctgtgactcc tcagcctctg ttttccccc 780
g 781

5
<210> 19
<211> 536
<212> DNA
<213> Homo sapiens

10
<400> 19
gcaagtgtgg gtggaggcca gtgcggggccc cacctgcccc ggggtcatcc ttgaacgccc 60
tgtgtggggc gagcagcctc agatgctgct gaagtgcaga cgcctccggg cctgaccctg 120
ggggcctgga gccacgctgg cagccctatg tgattaaacg ctggtgtccc caggccacgg 180
15 agcctggcag ggtccccaac ttcttgaacc cctgcttccc atctcagggg cgatggctcc 240
ccacgcttgg gagccttctg acccctgacc tgtgtcctct caccagcctc tccctgggct 300
ctgccctgag ctccctgggg cctgagcaag ttctctcccc gcccgcgcgc tccagcgtca 360
ctgggctgccc tgtctgctcg ccccggtgga ggggtgtctg tcccttccact gagggtccca 420
ccagccaggg ccacgaggtg caggccctgc ctgcccggcc acccacacgt cctaggaggg 480
20 ttggaggatg ccacctctgg cctcttctgg aacggagtct gattttggcc ccgcag 536

<210> 20
<211> 3179
<212> DNA
<213> Homo sapiens

25
<400> 20
atctcatggt tgaatcctaa tgtgcactgc atagacacca ctgtatgcaa ttacagaagc 60
ctgtgagtga acggggtggg ggtcagtgcg ggcccatggc ctggctgtgc atttacggaa 120
30 gtctatgagt gaatggggtt gtggtcagtg cgggccccatg gcctggctgg gcctgggagg 180
ttctctgatg tgtgaggcag gaggggaagg agggtagggg atagacagtg ggagccccc 240
ccctggaaga cataacagta agtccaggcc cgaaggcgag cagggatgct gggggcccg 300
cttgggcccgc ggggatgatg gagggcctgg ccagggtggc agggatgatg ggggccccag 360
ctgggggtggc aggggtgatg gggggggctg gtctgggtgg cggggaagat ggggaagcct 420
35 ggctgggccc cctcctcccc tgccctccac ctgcagccgt ggatccggat gtgcttcccc 480
gggtgcacatc ctctgggcca tcagctttca tggaggtggg gggcaggggc atgacacatc 540
cctgtataaa atccaggatt cctcctcctg aacgccccaa ctcaggttga aagtcacatt 600
ccgcctcttg ccattctctt aagagtagac caggattctg atctctgaag ggtgggtagg 660
gtggggcagt ggggggtgtg gacacaggag gcttcagggt ggggctgggt atgctctctc 720
40 atcctcttat catctcccag tctcatctct catcctctta tcatctccca gtctcatctg 780
tcttctctct atctcccagt ctcatctgtc atcctcttac catctcccag tctcatctct 840
tactctctta tctcctagtc tcattccagac ttacctccca gggcggtgct caggctcgca 900
gtggagcttg acatacgtcc ttctcaggc agaagggaact ggaaggattg cagagaacag 960
gagggggcgcc tcagagggac gcagctcttg ggtgaagaaa cagccccctc tcagaagtgt 1020
45 gcttgggcca caggaaccg agggccctgc gtgagtggtc ccagagcctt ccagcaggtc 1080
cctgggtggg ccttatggta tggccgggtc ctactgagtg cactctggac agggcttctg 1140
gtttgagtgc agcccggacg tgcctgggtg cggggtgggg gcttatggcc actggatatg 1200
gcgtcattta ttgctgtgct ttccagagaat gtctgagtgga ccgagcctaa tgtgtatggt 1260
gggcccagat ccacagactg tgcgttaaat gcaactctgt gcctggagcc cccgtatagg 1320
50 agctgtgagg aaggaggggc tcttggcagc cggcctgggg gcgcctttgc cctgcaaatc 1380
ggaaggggagc gggccccggc gccgtggggc gacgacctca agtgagagggt tggacagaa 1440
agggcgggga cttcccaggc gcagaggccg ctgctcaggc acacctgggt ttgaatcaca 1500
gaccaacagg tcaggccatt gttcagctat ccatcttcta caaagctcca gattcctgtt 1560
tctccgggtg ttttttgggt aaattttact caggattact tatatttttt gctaaagtat 1620
55 tagaccctta aaaaaggat ttgctttgat atggcttaac tactaagca cctactttat 1680
ttgtctgttt ttattttat ttattattat tattagagat ggtgtctact ctgtcacc 1740
gggtgttagt gcagtggcac agtcatggct cgctgtagcc gcaaaccctc aggtcgaagt 1800
gatectccgg cctcagcttc ccagagtgtc gggattacag gtgtgagcca ctgcccctgc 1860
ctggcacttt taaaaaccac tatgtaaggc cagggtccagt ggctccaca cctgtcatcc 1920
60 cagtgtttt ggaagccgag gcagaaggat tgtctgaggt caggagtttg agaccagcat 1980
gggttaacata gggagacccc atctctacaa aaaatgcaaa aagttatccg ggcgtgggg 2040
ccagcatctg tagtcccagc tgcctggggg gctgagtggt aggatcgctt gagcccgga 2100
gggtcatggt gcagtgtgct gtgattgtac catcgactc cagcctgggc aacagagtga 2160
gacctgtctt caaaaaaaaa aaaaaaaaaa gaaggagaag gagaagagaa gaagaaggaa 2220
65 gaaggaaaga gaagaagaag gaagaaggaa gaaagaaggaa gaaggaggcc tgcagggtgc 2280

18 / 18

taggttagact gtcaaattctc agagcaaaat gaaaataaca aagtttttaa gggaaagaaa 2340
aaccaccagct ctttggactt ccttaggcct gaacttcac tcaagcagct tccttccaca 2400
gacaagcgtg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggc 2460
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtgggtgttt tcctgcctca 2520
5 gccccacgct cctgccggtc ctgcacctgc tgtaaccgtc gatgttggtg ccagggtgcc 2580
acctgggaag gatgctgtgc agggggcttg ccaaactttg gtgggtttca gaagccccag 2640
gcactttgtg caggcacaat tacagccctt ccccaaagat gccacgtcc ttctcctgga 2700
acctgtgaat gtgtcaccog caaggcagag gctggtgaag gctgcagggt gaatcacggc 2760
tgccagtcag ccgatcttaa ggtcatcctg gattatctgg tgggcctgat atggccacaa 2820
10 gggcccttag aagtgaagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880
cactggccac tgctggcttt gagatggagg aggggggtccc cagccaagga atgggggcag 2940
ccgtccatg ctggaaaagc aagcaatcct ccccggtcct gagggcacac ggccctgccc 3000
acgcctcgat ttcaggccag tgggacctgt ttcagctttc cggcctccag agctgtaaga 3060
tgatgcgttt gtgttcagcc actaagctgc agtgattcgt cacagcagca aatggaatag 3120
15 cagtacaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179